ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

ABSOLUTELY CONVERGENT FOURIER EXPANSIONS FOR NON-COMMUTATIVE NORMED RINGS

BY S. BOCHNER AND R. S. PHILLIPS

(Received February 11, 1942)

I. The main theorem

The set R of elements $x, y, \cdots$ will be called a *normed ring* if

a) *R is a Banach space over the field of complex numbers.*

b) *The operation of multiplication is defined for elements of R. This operation possesses the usual algebraic properties.*

c) *R contains a unit element e.*

d) $\| xy \| \leqq \| x \| \cdot \| y \|, \qquad \| e \| = 1.$

Since commutativity of multiplication is not assumed we shall have to distinguish between left inverses, right inverses, and (two-sided) inverses. We note that a left inverse, if it exists, need not be unique.

Our principal result is the following theorem:

THEOREM 1. *If R' is the ring of periodic functions*

$$x(t) = \sum_{-\infty}^{\infty} a_n e^{int}, \qquad a_n \in R, \tag{1}$$

on $0 \leqq t < 2\pi$ to R, with

$$\sum \| a_n \| < \infty, \tag{2}$$

where the product $x(\cdot) \cdot y(\cdot)$ in R' is the function $x(t)y(t)$, then $x(t)$ has a left inverse in R' if $x(t_0)$ has a left inverse in R, for every t_0.

This is a generalization of the known theorem of N. Wiener[1] from complex numbers to a general R.

Denoting by $L(R)$ the Banach space of strongly integrable [Bochner, 2] functions $f(t)$ on $-\infty < t < \infty$ to R with the norm

$$\| f(\cdot) \| = \int_{-\infty}^{\infty} \| f(t) \| \, dt \tag{3}$$

we are able to obtain a generalization of another of Wiener's theorems.

THEOREM 2. *If $f(t) \in L(R)$ and if the Fourier transform*

$$x(u) = \int_{-\infty}^{\infty} f(t) e^{iut} \, dt \tag{4}$$

has a left inverse in R, for each real u, then the linear combinations

$$\sum_n A_n f(t - \lambda_n), \qquad A_n \in R, \tag{5}$$

[1] We shall quote from the memoir of Wiener [6] rather than from his book on the Fourier Integral.

are dense in $L(R)$. Conversely, if (5) *is dense in $L(R)$ then $x(u)$ has a left inverse throughout $(-\infty, \infty)$.*

The possibility of extending Wiener's results from numbers to ring elements was suggested to us by Gelfand's new method [3] of proving Wiener's theorem with the aid of maximal ideals in rings. It turns out that Wiener's own method leads very directly to our version of his theorem, whereas the adaptation of Gelfand's method to the non commutative case is slightly more elaborate. We shall present both methods. The normed ring approach has been found applicable to a slightly more general type of function ring.

II. Wiener's method

LEMMA 1. *If $a \in R$, $\| a \| < 1$, then the series*

$$c = e - a + a^2 - a^3 + \cdots$$

is absolutely convergent and

$$c(e + a) = (e + a)c = e. \tag{6}$$

Relation (6) can be verified by direct multiplication.

LEMMA 2. *If a has a left inverse a' (i.e. $a'a = e$) and if $\| b \| \cdot \| a' \| < 1$, then $a + b$ has a left inverse c, and*

$$c = a'(e - ba' + (ba')^2 - (ba')^3 + \cdots) \tag{7}$$

In fact, $a + b = (e + ba')a$, and hence, by Lemma 1,

$$c(a + b) = a'(e - ba' + (ba')^2 - \cdots)(e + ba')a = a'ea = e.$$

LEMMA 3. *The set U_l of all elements in R with a left inverse is an open set.*

PROOF. Lemma 2.

Since R is a Banach space, we first note that if $x(t)$ belongs to the R' of Theorem 1, it is in particular strongly integrable. Therefore series (1) is its Fourier series, and many familiar theorems hold. For instance, expansion (1) is unique, no matter how arrived at. In what follows, analysis of real and complex variables will be applied to the value space R without special emphasis.

LEMMA 4. *If $x(\cdot) \in R'$, and if $x(0)$ has a left inverse in R, then there exists an element*

$$y(t) = \sum c_n e^{int}$$

in R' with the following two properties:

(i) the coefficient c_0 has a left inverse c_0', and

$$\| c_0' \| \cdot \| \sum_{n=1}^{\infty} (c_n + c_{-n}) \| < 1;$$

(ii) in some interval $-\epsilon < t < \epsilon$, $y(t) = x(t)$.

PROOF. As in Wiener [6, p. 12, Lemma 2d], we introduced on the circle $-\pi \leqq t < \pi$ the numerical function

$$\omega_\epsilon(t) = \begin{cases} 1, & |t| < \epsilon \\ 2 - \dfrac{|t|}{\epsilon}, & \epsilon \leqq |t| < 2\epsilon \\ 0, & 2\epsilon \leqq |t|; \end{cases}$$

and the function

$$y_\epsilon(t) = \omega_\epsilon(t) \cdot x(t) + [1 - \omega_\epsilon(t)] \cdot x(0) = \sum c_n(\epsilon) e^{int}.$$

Obviously $y_\epsilon(t)$ satisfies property (ii) for each ϵ. As for property (i), Wiener's argument shows that

$$\lim_{\epsilon \to 0} c_0(\epsilon) = x(0),$$

$$\lim_{\epsilon \to 0} \sum_{n=1} \| c_n(\epsilon) + c_{-n}(\epsilon) \| = 0.$$

Thus by Lemma 2 property (i) holds for sufficiently small ϵ.

LEMMA 5. *If $y(\cdot) \in R'$, and $y(t)$ satisfies property (i) of Lemma 4, then $y(\cdot)$ has a left inverse $y'(\cdot)$ in R'.*

In fact, putting

$$B(t) = \sum_{n=1} (c_n e^{int} + c_{-n} e^{-int}),$$

we have by Lemma 2,

$$y'(t) = c_0'[e - (B(t)c_0' + (B(t)c_0')^2 - (B(t)c_0')^3 + \cdots]$$

Now if $x(\cdot) \in R'$ and if for all t, $x(t) \in U_l$, then by Lemmas 4 and 5 there exists for each t_0 a function $y_{t_0}(\cdot) \in R'$ such that $y_{t_0}(t) \cdot x(t) = e$ in some interval $(t_0 - \epsilon, t_0 + \epsilon)$. We can now piece together a finite number of these functions $y_{t_0}(\cdot)$ and obtain a function $x'(\cdot) \in R'$ for which $x'(t) \cdot x(t) = e$ for all t. For details of the argument see Wiener [6, pp. 10–11, Lemma II_b]. This concludes the proof of Theorem 1.

We now turn to the proof of Theorem 2. Our main step is

LEMMA 6. *Given constants $-\pi < \alpha < a < b < \beta < \pi$, if $x_1(\cdot)$, $x_2(\cdot)$ belong to R' where $x_2(u)$ has a left inverse for $\alpha < u < \beta$, and $x_1(u)$ vanishes outside $a \leqq u \leqq b$, then there exists an element $x_3(\cdot)$ of R' which vanishes outside $\alpha \leqq u \leqq \beta$ such that*

$$x_1(u) = x_3(u)x_2(u) \tag{8}$$

in $-\pi \leqq u < \pi$.

PROOF. By Lemmas 4 and 5 corresponding to any u_0 in $a \leqq u \leqq b$, there exists a function $y_{u_0}(\cdot) \in R'$ such that in some interval $(u_0 - \epsilon, u_0 + \epsilon)$, $y_{u_0}(u)x_2(u) = e$. As above, we can define a function $x_2'(u)$ which belongs to R'

and for which $x_2'(u)x_2(u) = e$ in $a \leqq u \leqq b$. Let $x_3(u) = x_1(u) \cdot x_2'(u)$. Then $x_3(u) = 0$ in the intervals $-\pi \leqq u \leqq \alpha, \beta \leqq u < \pi$. In addition

$$x_3(u) \in R', \quad \text{and} \quad x_1(u) = x_3(u)x_2(u).$$

LEMMA 7. *If $x(u)$ is strongly integrable in $(-\pi, \pi)$ and vanishes in $(-\pi, -\pi + \epsilon)$ and $(\pi - \epsilon, \pi)$, and if we put*

$$f(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} x(u)e^{-iut}\,du, \qquad a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} x(u)e^{-inu}\,du$$

then $\int_{-\infty}^{\infty} \| f(t) \|\, dt < \infty$ if and only if $\sum \| a_n \| < \infty$.

PROOF. As in Wiener [6, p. 14–15, Lemma IIf].

From Lemmas 6, 7 we now conclude

LEMMA 8. *If $g(t)$ and $f(t)$ both belong to $L(R)$, if $x_1(u) = \int_{-\infty}^{\infty} g(t)e^{-iut}\,dt$ vanishes outside some interval (a, b), and if $x_2(u) = \int_{-\infty}^{\infty} f(t)e^{-iut}\,dt$ vanishes outside some larger interval (α, β), $\alpha < a$, $b < \beta$, $x_2(u)$ having a left inverse for each u in $\alpha < u < \beta$, then there exists an element $h(t)$ in $L(R)$ such that*

$$g(t) = \int_{-\infty}^{\infty} h(\lambda)f(t - \lambda)\,d\lambda. \tag{9}$$

Integral (9) can be approximated in $L(R)$ by sums of the form (5). The transition from Lemma 8 to Theorem 2 can be effected by Fejer approximation as in Wiener [6, p. 16–18].

III. Irreducibility and maximal ideals

The following lemmas are the algebraic basis for an adaptation of Gelfand's methods.

Let $V = (a, b, c, \cdots)$ be a Banach space over the complex numbers, and let $A = (\alpha, \beta, \cdots)$ be the ring of all bounded linear transformations on V to V. Then $\alpha(a + b) = \alpha a + \beta b$, $(\beta\alpha)a = \beta(\alpha a)$. If V_0 is a subset of V and A_0 a subset of A then A_0V_0 will denote the set of all elements (αa) of V for $\alpha \in A_0$ and $a \in A_0$.

V will be said to be *irreducible* over the subring A_0 of A if for every $a \in V$, $(a \neq 0)$, $A_0a = V$. This means that V has no subspaces invariant with respect to A_0.

LEMMA 9. *If V is irreducible over A_0, and if A' denotes the set of all elements of A which commute with all elements of A_0, then A' is isomorphic to the field of complex numbers.*

In fact if $\alpha', \beta' \in A'$ and $\alpha_0 \in A_0$, then $\alpha'\alpha_0 = \alpha_0\alpha'$, $\beta'\alpha_0 = \alpha_0\beta'$ implies $(\alpha' \pm \beta')\alpha_0 = \alpha_0(\alpha' \pm \beta')$. Similarly $\alpha'\beta'\alpha_0 = \alpha'\alpha_0\beta' = \alpha_0\alpha'\beta'$. Thus A' is a ring. Clearly A' contains the null element and the identity. Now, if $\alpha' \in A'$, $\alpha' \neq 0$, consider the set $W = \alpha'V$. The irreducibility implies $A_0W = A_0V = V$,

and the commutativity implies $\alpha' A_0 = A_0 \alpha'$. Therefore $W = \alpha' V = \alpha'(A_0 V) = A_0(\alpha' V) = A_0 W = V$. That is, the range of the homomorphism α' is all of V Now suppose that for some $a \in V$, $a \neq 0$, we had $\alpha' a = 0$. Then $\alpha' V = \alpha'(A_0 a) = A_0(\alpha' a) = 0$ which is contrary to $a' V = V$. α' is therefore a one to one linear bounded transformation on V to V. By a theorem due to Banach (1, p. 41, Theorem 5), the inverse β' is a linear bounded transformation and hence belongs to A. $\beta' \alpha_0 = \beta' \alpha_0 \alpha' \beta' = \beta' \alpha' \alpha_0 \beta' = \alpha_0 \beta'$ for all $\alpha_0 \in A_0$. Thus β' belongs to A'. A' is therefore a field. Clearly a limit (in A) of transformations which commute with elements of A_0 also commutes with these elements. A' is thus a complete normed field over the complex numbers. Such a field is isomorphic to the field of complex numbers. Since the proof of this is the same in the commutative and noncommutative cases we refer the reader to Gelfand [3, pp. 6–8].

If R is a ring with unit element, a *left ideal* I is a subset with the properties:

1) *if $x \in I$, $y \in I$, then $px + qy \in I$ for any elements p, q from R.*

2) *I is a proper subset of R.*

A *maximal* left ideal is one not contained in a larger left ideal.

LEMMA 10. *If R is a ring with unit, if I is a maximal left ideal, if V is the addition group of the cosets R/I, and if A_0 is the ring of homomorphisms of V onto itself as produced by multiplying V by elements of R from the left, then V is irreducible with respect to A_0.*

If $x \in R$, then the element of A_0 corresponding to x will be denoted by $\mu(x)$. For fixed $a \in V$, $a \neq 0$, we form the set $V_0 = \mu(R)a$. Since R is a linear space, V_0 is a linear subspace of V. Also, since R contains a unit element e, $\mu(e)a = a \neq 0$. Thus V_0 contains the null element and some other elements. Now form the set S of all elements of R contained in the cosets composing V_0. Then S includes I as a proper part. On the other hand, S is a left ideal. As I was supposed maximal, $S = R$. Thus $V_0 = V$; that is $A_0 a = V$.

LEMMA 11. *If R, I, V, A_0 have the same meaning as in Lemma 10, and if for fixed $x \in R$ and every maximal left ideal I the corresponding element $\mu(x)$ of A_0 has a left inverse in A_0, then x has a left inverse in R.*

PROOF. Since R contains a unit element e, every left ideal is contained in a maximal left ideal, and consequently an element x of R has a left inverse x' if (and only if) it is contained in no maximal ideal. See Gelfand [3, p. 8–9].

If $x'x = e$, then $\mu(x')\mu(x) = \mu(e)$. Denoting by a_e and a_0 the elements of V which as cosets in R/I contain the elements e and 0 respectively, we have in particular

$$\mu(x')\mu(x)a_e = \mu(e)a_e . \tag{10}$$

Since $e \cdot e = e$, we have $\mu(e)a_e = a_e$. On the other hand, $\mu(x)a_e$ contains x. Hence if x were contained in $a_0 = I$, we would have $\mu(x)a_e = a_0$, and consequently $\mu(x')\mu(x)a_e = a_0$. This completes the proof of Lemma 11.

Finally we have to make several statements for *normed* rings.

LEMMA 12. *Every maximal ideal of a normed ring is closed.*

For the proof see Gelfand [3, p. 8].

Preparatory to an amendment of Lemma 11, we first observe that if R is normed, and X is a closed linear subspace, then R/X is made into a Banach space by introducing the norm

$$|| Y || = \inf_{y \in Y} || y ||$$

for any Y of R/X [see 3].

LEMMA 13. *If R, I, V, A_0, have the same meaning as in Lemma 10, then V is a Banach space, A_0 is a normed ring, and for the norm $|| \mu(x) ||$ in A_0 we have*

$$|| \mu(x) || \leqq || x ||, \qquad x \in R.$$

In fact,

$$|| \mu(x) || = \sup_{a \in V} \frac{|| \mu(x)\, a ||}{|| a ||} = \sup_{a \in V} \frac{\inf [|| xy || \mid y \in a]}{\inf [|| y || \mid y \in a]} \leqq || x ||.$$

IV. Theorem 3

ASSUMPTIONS. *Let F denote a (commutative) ring of complex valued functions $f(t)$ on a point set $[t]$. Ring multiplication is ordinary multiplication $f(t) \cdot g(t)$, and the constant function $f(t) \equiv 1$ belongs to F and is its unit. The norm in F will be denoted by ordinary bars: $| f(t) |$. By $M(f)$ we shall denote any continuous ring homomorphism from F to complex numbers. Thus in particular $M(fg) = M(f) \cdot M(g)$.*

Let $R = (x, y, \cdots)$ denote any (non-commutative) normed ring with unit e and norm $|| x ||$, and let R' denote a family of functions $x(\cdot) = x(t)$ from $[t]$ to R with the following properties:

1) *R' is a ring under point multiplication $x(t)y(t)$.*

2) *If $x_1, \cdots, x_n \in R$, $f_1(t) \cdots, f_n(t) \in F$, then*

$$x_1 f_1(\cdot) + \cdots + x_n f_n(\cdot) \tag{11}$$

belongs to R'. If $x \in R$, $f(t) \in F$, then the element $xf(\cdot)$ of R' will be denoted by x^f.

3) *R' is a normed ring with norm $|| x(\cdot) ||$ for which*

$$|| x^f || = || x || \cdot | f |.$$

4) *The linear combinations (11) are dense in R'.*

5) *If $x(\cdot) = x_1^{f_1} + \cdots + x_n^{f_n}$, then for every homomorphism $M(f)$ of F,*

$$|| x_1 M(f_1) + \cdots + x_n M(f_n) || \leqq || x (\cdot) ||.$$

On the basis of 4) and 5) every $M(f)$ gives rise to a continuous homomorphism $\mathbf{M}(x(\cdot))$ from R' to R, with the property

$$\mathbf{M}(x^f) = x M(f).$$

We will call $\mathbf{M}$ *a* generated *homomorphism.*

CONCLUSION. *An element $x(\cdot) \in R'$ has a left inverse if for every generated homomorphism $\mathbf{M}$, the element $\mathbf{M}(x(\cdot))$ of R has a left inverse in R.*

PROOF. Replacing R by R' in Lemma 11, we consider an arbitrary maximal left ideal I and for any element $x(\cdot) \in R'$ the corresponding element $\mu(x(\cdot))$ of A_0. The elements $e^f(\cdot) = ef(\cdot)$ of R', $e =$ unit in R, $f \in F$, are commutative with any element x^g of R' and hence by 4) with every element of R'. Therefore, by Lemmas 9, 10, and 13, $\mu(e^f(\cdot))$ can be looked upon as a continuous ring homomorphism from F to complex numbers. More precisely there exists an $M(f)$ such that

$$\mu(e^f) = e_1 M(f) \qquad (e_1 = \mu(e \cdot 1) \text{ is unit of } A).$$

For an element of the form (11) we have

$$\mu(x(\cdot)) = \mu(x_1^{f_1} + \cdots + x_k^{f_k}) = \mu(x_1 \cdot 1)\mu(e_{f_1}) + \cdots + \mu(x_k \cdot 1)\mu(e_{f_k})$$
$$= \mu(x_1 \cdot 1)M(f_1) + \cdots + \mu(x_k \cdot 1)M(f_k) = \mu([x_1 M(f_1) + \cdots + x_k M(f_k)] \cdot 1)$$

and thus

$$\mu(x(\cdot)) = \mu(\mathbf{M}(x(\cdot)) \cdot 1) \tag{12}$$

By property 5) in conjunction with Lemma 13 this relation is valid for all elements $x(\cdot) \in R'$. Now if $\mathbf{M}(x(t))$ has a left inverse x' in R, then

$$\mu(x' \cdot 1) \cdot \mu(x(\cdot)) = \mu(x' \cdot x(\cdot)).$$

Applying (12) to $x' \cdot x(\cdot)$ instead of $x(\cdot)$, this is

$$= \mu(\mathbf{M}(x' \cdot (x(\cdot)) \cdot 1) = \mu(x' \cdot \mathbf{M}(x(\cdot)) \cdot 1) = \mu(e \cdot 1) = e_1 .$$

Thus $\mu(x(\cdot))$ has a left inverse, and by Lemma 11, $x(\cdot)$ has a left inverse in R'.

A PARTICULAR ASSUMPTION. Corresponding to any $M(f)$ there exists a point t_0 such that

$$M(f) = f(t_0) \qquad f \in F.$$

In this case, property 5) can be replaced by the simpler property: $\| x(t_0) \| \leqq \| x(\cdot) \|$, for each t_0.

A PARTICULAR CONCLUSION. *The element $x(\cdot) \in R'$ has a left inverse in R', provided $x(t)$ has a left inverse in R, for all t.*

The proof is obvious.

Adjunction of unit. If the ring F has no unit, we make a formal adjunction of a unit 1, and we consider the enlarged ring $\bar{F}$: $\lambda 1 + f$, with the norm $|\lambda| + |f|$ ($\lambda =$ complex number). This leads to the ring $\bar{R}'$ of elements $\bar{x}(\cdot) = \lambda e \cdot 1 + x(\cdot)$ where $\| \bar{x}(\cdot) \| = |\lambda| + \| x(\cdot) \|$. It can be shown, Gelfand [3], that every homomorphism $\bar{M} = \bar{M}(\lambda 1 + f)$ to the complex numbers is either $\lambda + M(f)$ where M is a homomorphism of F, or it is the exceptional homomorphism $\bar{M}(\lambda 1 + f) = \lambda$. We can then obtain the following *conclusion*: If F has no unit, then the element $\bar{x}(\cdot) = \lambda e \cdot 1 + x(\cdot)$ with $\lambda \neq 0$ and $x(\cdot) \in R'$, has a left inverse of the form $\lambda' e \cdot 1 + x'(\cdot)$, if for every generated homomorphism $\mathbf{M}$ the element $\lambda e + \mathbf{M}(x(\cdot)) \in R$ has a left inverse in R. In the particular circumstances cited before, it is again sufficient that $\lambda e + x(t)$ shall have an inverse for each t.

V. Expansions on groups. Applications

By known results on topological groups, and by results on numerical functions by Gelfand [4], and Gelfand-Rykov [5], Theorem 3 leads easily to the following theorems.

THEOREM 4. *If $\Gamma = (\alpha, \beta, \cdots)$ is a commutative group of addition with discrete topology and $G = (t, s, \cdots)$ is the compact dual group of addition; if $\{\chi(\alpha, t)\}$, $\alpha \in \Gamma$, $t \in G$ are the characters from Γ to G; if R is any normed ring; and if R' is the ring of functions on G,*

$$x(t) = \sum_\alpha a_\alpha \chi(\alpha, t) \qquad a_\alpha \in R \tag{13}$$

$$\sum_\alpha \| a_\alpha \| < \infty, \tag{14}$$

then $x(\cdot)$ has a left inverse in R', provided $x(t)$ has a left inverse in R, for all t.

More generally, if (14) is replaced by

$$\sum_\alpha e^{q_\alpha} \| a_\alpha \| < \infty,$$

where the real numbers q_α have the properties $q_{\alpha+\beta} \leqq q_\alpha + q_\beta$ and $q_0 = 0$, then the condition is that

$$\sum_\alpha a_\alpha e^{p_\alpha} \chi(\alpha, t)$$

shall have a left inverse in R for all t, and every system of real numbers p_α, for which $p_{\alpha+\beta} = p_\alpha + p_\beta$, $p_0 = 0$, $p_\alpha \leqq q_\alpha$.

However, if Γ is a locally bicompact group with unique Haar measure $d\alpha$; if $\chi(\alpha, t)$ are continuous characters; if R' is formed by

$$x(t) = \int_\Gamma a_\alpha \chi(\alpha, t)\, d\alpha, \qquad a_\alpha \in R$$

where

$$\int_\Gamma e^{q_\alpha} \| a_\alpha \|\, d\alpha < \infty;$$

and if R' has no unit, but 1 is an adjoined unit, then the element $\bar{x}(\cdot) = \lambda 1 + x(\cdot)$ has a left inverse of the form $\lambda' 1 + x'(\cdot)$ provided

$$\lambda \cdot \left[\lambda \cdot e + \int_\Gamma a_\alpha e^{p_\alpha} \chi(\alpha, t)\, d\alpha \right]$$

has a left inverse for all t, and every continuous system p_α with the previous properties.

In the case of a discrete group Γ it is often natural to consider not the total compact group G but a dense subgroup G_0. For instance, if Γ is the linear group $-\infty < \alpha < \infty$ without topology then it is natural to consider the characters $\chi(\alpha, t) = e^{i\alpha t}$ with $-\infty < t < \infty$, dense in the total group G. It is not

necessary to make sure of the existence of $x'(t)$ for all G, but it is sufficient to know that the left inverse $x'(t)$ exists and is bounded in norm $(-\infty, \infty)$. This results from the following lemma:

LEMMA 14. *Suppose $x_n \to x$ in R. If x_n' is a left inverse of x_n and the $\| x_n' \|$ are bounded, then x possesses a left inverse.*

PROOF. $e - x_n'x = e - x_n'x_n + x_n'x_n - x_n'x$. Hence $\| e - x_n'x \| \leqq \| x_n' \| \cdot \| x_n - x \| \to \infty$. By Lemma 1 there exists for sufficiently large n, an inverse y_n of $x_n'x$ so that $y_n y_n'$ is a left inverse for x

However, if Γ is the continuous group $-\infty < \alpha < \infty$ with the ordinary Lebesque measure as measure, the line $(-\infty, \infty)$ is the complete group G. Thus we obtain the following:

THEOREM 5. *If $x(t) = \sum a_n e^{i\alpha_n t}$, $a_n \in R$, $\sum \| a_n \| < \infty$, and if a left inverse $y(t)$ exists for each t and $\| y(t) \| < C$, then there exists a function $x'(t) = \sum b_n e^{i\beta_n t}$ with $\sum \| b_n \| < \infty$ such that $x'(t) \cdot x(t) = e$.*

If $a(\alpha) \in R$, $-\infty < \alpha < \infty$, and $a(\alpha)$ is strongly integrable $\int_{-\infty}^{\infty} \| a(\alpha) \| \, d\alpha < \infty$, and if for complex $\lambda \neq 0$, $\lambda e + \int_{-\infty}^{\infty} a(\alpha) e^{i\alpha t} \, d\alpha$ has a left inverse for all t, then there exists a left inverse of the form $\lambda' e + \int_{-\infty}^{\infty} b(\alpha) e^{i\alpha t} \, d\alpha$, with $\int_{-\infty}^{\infty} \| b(\alpha) \| d\alpha < \infty$.

Difference and integral equations. Now, consider functions $\varphi(t)$, $\psi(t)$, $\cdots$ on the line $-\infty < t < \infty$ with values in a space on which R operates; and assume that they form a Banach space B, and that the norm in B is invariant under translation. Then, all symbols having the same meaning as in Theorem 5, we see that the difference equation

$$\sum a_n \varphi(t - \alpha_n) = \psi(t) \tag{15}$$

has a solution

$$\varphi(t) = \sum b_n \psi(t - \beta_n), \tag{16}$$

and the integral equation

$$\varphi(t) + \int_{-\infty}^{\infty} a(\alpha) \varphi(t - \alpha) \, d\alpha = \psi(t)$$

has a solution

$$\varphi(t) = \psi(t) + \int_{-\infty}^{\infty} b(\beta) \psi(t - \beta) \, d\beta.$$

For instance, let R be the space of k-dimensional matrices of complex numbers

$$x \equiv (x_{ij}) \qquad i, j = 1, \cdots, k,$$

and let $\| x \|$ be the maximum of $\sum_{i,j=1}^{k} x_{ij} p_i q_j$ for $\sum | p_i |^2 \leqq 1$, $\sum | q_j |^2 \leqq 1$. Then we obtain the following result.

If $a_n = (a_{ij}^n)$, $\sum \| a_n \| < \infty$, $\alpha_n =$ real, and if for

$$D(t) = \left(\sum_{n=1}^{\infty} a_{ij}^n e^{i\alpha_n t} \right) \qquad i, j = 1, \cdots, k,$$

$$\| D^{-1}(t) \| \leqq C < \infty \qquad -\infty < t < \infty,$$

then there exist matrices $b_n = (b_{ij}^n)$, $\sum \| b_n \| < \infty$, and real numbers β_n, such that the system of equations

$$\sum_{n=1}^{\infty} \sum_{j=1}^{k} a_{ij}^n \varphi_j(t + \alpha_n) = \psi_i(t) \qquad i = 1, \cdots, k$$

has a solution of the form

$$\varphi_i(t) = \sum_{n=1}^{\infty} \sum_{j=1}^{k} b_{ij}^n \psi_j(t + \beta_n).$$

If the functions $\psi_j(t)$ belong to Lebesgue class L_p, to B, M, C etc., then the functions $\varphi_i(t)$ belong to the same class respectively.

PRINCETON UNIVERSITY,
HARVARD UNIVERSITY

REFERENCES

1. S. BANACH, *Théorie des Operations Linéaires*, Warsaw, 1932.
2. S. BOCHNER, *Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind*, Fund. Math., vol. 20, 1933, pp. 262–276.
3. I. GELFAND, *Normiete Ringe*, Recueil Mathématique, vol. 9 (51), 1941, pp. 3–24.
4. I. GELFAND, *Über absolut konvergente trigonometrische Reihen und Integrale*, Recueil Mathématique, vol. 9 (51), 1941, pp. 51–66.
5. I. GELFAND AND D. RAIKOV, *On the Theory of Characters of Commutative Topological Groups*, Comptes Rendus de l'Adademie des Sciences de l'URSS, vol. 28, 1940, pp. 195–6.
6. N. WIENER, *Tauberian theorems*, these Annals, vol. 33, 1932, pp. 1–100.

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

ON THE LAW OF THE ITERATED LOGARITHM

BY PAUL ERDÖS

(Received December 26, 1941)

Introduction

Let t be a real number $(0 \leqq t \leqq 1)$, and let $t = 0.\epsilon_1(t)\epsilon_2(t) \cdots$ be its dyadic expansion, or equivalently,

$$t = \frac{\epsilon_1(t)}{2} + \frac{\epsilon_2(t)}{2^2} + \cdots + \frac{\epsilon_n(t)}{2^n} + \cdots, \tag{0.1}$$

where $\epsilon_n(t) = 0$ or 1 according as the integral part of $2^n t$ is even or odd. It is well known that $\{\epsilon_n(t)\}$ $(n = 1, 2, \cdots)$ is an independent system in the sense of probability,[1] and that

$$\int_0^1 \epsilon_n(t)\, dt = \frac{1}{2}, \qquad \int_0^1 \left(\epsilon_n(t) - \frac{1}{2}\right)^2 dt = \frac{1}{4}. \tag{0.2}$$

Let us further put

$$f_n(t) = \sum_{k=1}^{n} \epsilon_k(t) - \frac{n}{2}. \tag{0.3}$$

It was proved by A. Khintchine[2] and A. Kolmogoroff[3] that

$$\limsup_{n\to\infty} \frac{f_n(t)}{\left(\frac{n}{2}\log\log n\right)^{\frac{1}{2}}} = 1 \tag{0.4}$$

for almost all t.

Let $\varphi(n)$ be a monotone increasing non-negative function defined for all sufficiently large integers. Following P. Lévy we say that $\varphi(n)$ belongs to the *upper class* if, for almost all t, there exist only finitely many n such that

$$f_n(t) > \varphi(n); \tag{0.5}$$

and $\varphi(n)$ belongs to the *lower class* if, for almost all t, there exist infinitely many n such that (0.5) is true. According to the well-known law of 0 or 1, each $\varphi(n)$ must belong to one of these classes. Then the result of A. Khintchine and A. Kolmogoroff stated above means that $\varphi(n) = (1 + \epsilon)(\frac{1}{2}n \operatorname{loglog} n)^{\frac{1}{2}}$ belongs to the upper class if $\epsilon > 0$, and to the lower class if $\epsilon < 0$.

The purpose of the present paper is to give a sharpening of this result. The

[1] Cf. M. Kac and H. Steinhaus, *Sur les fonctions indépendentes*, Studia Math. 6 (1936), 46–58, 59–66, 89–97.

[2] A. Khintchine, *Asymptotische Gesetz der Wahrscheinlichkeitsrechnung*, Berlin, 1933.

[3] A. Kolmogoroff, *Über das Gesetz der iterierten Logarithmus*, Math. Annalen, 101 (1929), 126–135.

main results are stated in Theorems 1, 2, 3, 4, and 5 below. Among other results, it follows from Theorem 3 that, for $k > 3$,

$$\varphi(n) = \left(\frac{n}{2 \log\log n}\right)^{\frac{1}{2}} \left(\log\log n + \frac{3}{4}\log_3 n + \frac{1}{2}\log_4 n + \cdots + \frac{1}{2}\log_{k-1} n + \left(\frac{1}{2} + \epsilon\right)\log_k n\right) \tag{0.6}$$

belongs to the upper class if $\epsilon > 0$ and to the lower class if $\epsilon \leqq 0$.

Our proof is direct and elementary. We do not assume the result of A. Khintchine and A. Kolmogoroff, and the paper can be read without knowledge of any particular results concerning the law of the iterated logarithm. The only facts we need are the notion of independence, and the well known inequality

$$c_1 \frac{n}{x} e^{-2x^2/n} < Pr(A_n(x)) < c_2 \frac{n}{x} e^{-2x^2/n}, \tag{0.7}$$

where

$$A_n(x) = E[t: f_n(t) > x] \tag{0.8}$$

means the set of all real numbers t $(0 \leqq t \leqq 1)$ satisfying $f_n(t) > x$, and $Pr(A)$ means the ordinary Lebesque measure of a measurable set A in the interval $0 \leqq t \leqq 1$. c_i $(i = 1, 2, \cdots)$ will denote positive constants.

Throughout the present paper, the sequence $\{m_n\}$ $(n = 1, 2, \cdots)$ defined by $m_1 = 1$ and

$$m_n = [e^{n/\log n}], \qquad n = 2, 3, \cdots, \tag{0.9}$$

will play a fundamental rôle. The fact that we adopt the sequence $\{m_n\}$ $(n = 1, 2, \cdots)$ instead of $\{a^n\}$ $(n = 1, 2, \cdots)$, which was used by A. Khintchine and A. Kolmogoroff, is essential in our proof, and will enable us to obtain our sharper results. The following inequalities, which are easy to prove, will be used very often:

$$m_n < m_{n+1} < c_3 m_n, \tag{0.10}$$

$$c_4 \frac{m_n}{\log\log m_n} < m_{n+1} - m_n < c_5 \frac{m_n}{\log\log m_n} \tag{0.11}$$

$$c_6 \left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}} < (m_{n+1} \log\log m_{n+1})^{\frac{1}{2}} - (m_n \log\log m_n)^{\frac{1}{2}} < c_7 \left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}}. \tag{0.12}$$

It is not difficult to extend our results to the case in which the parameter n is continuous, i.e. the case of Brownian motion.[4] We can define the upper and

[4] Cf. A. Khintchine, loc. cit. 2. Cf. also N. Wiener, *Differential space*, Journal of Math. and Phys. 2 (1923), 131–174, and the book of P. Lévy quoted in footnote 5.

the lower classes in this case, and can obtain the corresponding results. It was stated by P. Lévy[5] that A. Kolmogoroff has proved the following result: Let $\psi(\lambda) = \varphi(\lambda)/\lambda^{\frac{1}{2}}$ be monotone increasing. Then a necessary and sufficient condition that $\varphi(\lambda)$ belong to the lower class is given by the divergence of the integral

$$\int_0^\infty \psi(\lambda) e^{-2(\psi(\lambda))^2} \frac{d\lambda}{\lambda}. \tag{0.13}$$

It is easy to see that this is equivalent to Theorem 4. As far as I know, the proof of A. Kolmogoroff has not been published. Recently, J. Ville[6] proved that the divergence of (0.13) is necessary. This corresponds to a special case of Theorem 1, but his proof is entirely different from ours.

1

THEOREM 1. *$\varphi(n)$ belongs to the upper class if it is monotone increasing and if*

$$\sum_{n=1}^{\infty} Pr(A_{m_n}(\varphi(m_n))) < \infty. \tag{1.1}$$

PROOF. First we remark that we may assume that

$$\varphi(n) \leqq (n \operatorname{loglog} n)^{\frac{1}{2}} \tag{1.2}$$

for sufficiently large n. Indeed, otherwise we may consider $\varphi_1(n) = \min(\varphi(n), (n \operatorname{loglog} n)^{\frac{1}{2}})$ instead of $\varphi(n)$. It is clear that $\varphi_1(n)$ is monotone increasing, that $\varphi_1(n)$ satisfies (1.1) if $\varphi(n)$ does (because, by (0.7), $\varphi_0(n) = (n \log\log n)^{\frac{1}{2}}$ satisfies (1.1)); and that if $\varphi_1(n)$ belongs to the upper class so does $\varphi(n)$ too.

Next we notice that, under the assumption (1.2), we have

$$Pr(A_{m_{n+1}}(\varphi(m_n))) < c_8 Pr(A_{m_n}(\varphi(m_n))). \tag{1.3}$$

This is an easy consequence of the relations (0.7), (0.10) and (0.11). We omit the proof.

Now assume that Theorem 1 is not true. Then there exists a constant $c_9 > 0$ such that, for any $M_0 = m_{n_0}$, there exists an $N_0 = m_{n_0'}$ $(n_0' > n_0)$ such that

$$Pr\Big(\sum_{M_0 < u \leqq N_0} A_u(\varphi(u))\Big) > c_9 > 0. \tag{1.4}$$

Let us put

$$\begin{aligned} B(u) &= A_u(\varphi(u)) - A_u(\varphi(u)) \sum_{M_0 < v < u} A_v(\varphi(v)) \\ &= E[t: f_u(t) > \varphi(u); f_v(t) \leqq \varphi(v), M_0 < v < u]. \end{aligned} \tag{1.5}$$

[5] P. Lévy, *Théorie de l'addition des variables aléatoires*, Paris, 1937.
[6] J. Ville, *Étude critique de la notion de collectif*, Paris, 1937.

Then $\{B(u)\}$ $(M_0 < u \leqq N_0)$ are mutually disjoint, and

$$\sum_{M_0 < u \leqq N_0} B(u) = \sum_{M_0 < u \leqq N_0} A_u(\varphi(u)). \tag{1.6}$$

For each u $(M_0 < u \leqq N_0)$ take an n $(n_0 \leqq n < n_0')$ such that $m_n < u \leqq m_{n+1}$, and put

$$\Delta^+_{u,m_{n+1}} = E[t: f_{m_{n+1}}(t) - f_u(t) \geqq 0]. \tag{1.7}$$

Then it is clear that $B(u)$ and $\Delta^+_{u,m_{n+1}}$ are independent, and hence

$$Pr(B(u)\cdot\Delta^+_{u,m_{n+1}}) = Pr(B(u))Pr(\Delta^+_{u,m_{n+1}}) \geqq \tfrac{1}{2}\cdot Pr(B(u)). \tag{1.8}$$

On the other hand, since $t \in B(u)\cdot\Delta^+_{u,m_{n+1}}$ implies $f_{m_{n+1}}(t) \geqq f_u(t) > \varphi(u) \geqq \varphi(m_n)$, we have

$$B(u)\cdot\Delta^+_{u,m_{n+1}} \subset A_{m_{n+1}}(\varphi(m_n)) \tag{1.9}$$

for $m_n < u \leqq m_{n+1}$. Hence, since $\{B(u)\cdot\Delta^+_{u,m_{n+1}}\}$ $(M_0 < u \leqq N_0)$ are mutually disjoint, we have, by (1.3), (1.8), (1.6) and (1.4),

$$\begin{aligned} c_8 \sum_{M_0 < u \leqq N_0} Pr(A_{m_n}(\varphi(m_n))) &\geqq \sum_{M_0 < u \leqq N_0} Pr(A_{m_{n+1}}(\varphi(m_n))) \\ &\geqq \sum_{M_0 < u \leqq N_0} Pr(B(u)\cdot\Delta^+_{u,m_{n+1}}) \geqq \tfrac{1}{2} \sum_{M_0 < u \leqq N_0} Pr(B(u)) \\ &\geqq \tfrac{1}{2} Pr\Big(\sum_{M_0 < u \leqq N_0} B(u)\Big) = \tfrac{1}{2} Pr\Big(\sum_{M_0 < u \leqq N_0} A_u(\varphi(u))\Big) > \frac{c_9}{2} > 0. \end{aligned} \tag{1.10}$$

Since c_8 and c_9 are positive constants, and since $M_0 = m_{n_0}$ can be arbitrarily large, this contradicts to the assumption (1.1). This proves Theorem 1.

COROLLARY 1. *$\varphi(n) = (1/(2)^{\frac{1}{2}} + \epsilon)$ $(n \log\log n)^{\frac{1}{2}}$ belongs to the upper class for $\epsilon > 0$.*

COROLLARY 2. *The expression* (0.6) *belongs to the upper class for $\epsilon > 0$.*

PROOF. Follows immediately from Theorem 1 and (0.7).

2

THEOREM 2. *If $\varphi(n)$ is monotone increasing, then a necessary and sufficient condition that $\varphi(n)$ belong to the lower class is that, for almost all t, there exist infinitely many n such that*

$$f_{m_n}(t) > \varphi(m_n). \tag{2.1}$$

PROOF. The sufficiency is obvious. In order to prove the necessity, let us assume that $\varphi(n)$ belongs to the lower class. First we remark that we may assume

$$\varphi(n) \leqq (n \log\log n)^{\frac{1}{2}} \tag{2.2}$$

for sufficiently large n. Indeed, by Corollary 1 to Theorem 1, $\varphi_0(n) = (n \log\log n)^{\frac{1}{2}}$ belongs to the upper class. Hence, if we put $\varphi_1(n) = \min(\varphi(n), \varphi_0(n))$, then

$\varphi_1(n)$ belongs to the lower class if $\varphi(n)$ does; and if the necessity of the condition is proved for $\varphi_1(n)$, then it is obviously true for $\varphi(n)$ too.

By assumption, there exists a constant $c_{10} > 0$ such that for any $M_0 = m_{n_0}$ there exists an $N_0 = m_{n_0'}$ $(n_0' > n_0)$ such that

$$Pr\Big(\sum_{M_0 < u \leqq N_0} A_u(\varphi(u))\Big) > c_{10}. \tag{2.3}$$

Let us put

$$\begin{aligned} C(u) &= A_u(\varphi(u)) - A_u(\varphi(u)) \cdot \sum_{u < v \leqq N_0} A_v(\varphi(v)) \\ &= E[t: f_u(t) > \varphi(u); f_v(t) \leqq \varphi(v),\ u < v \leqq N_0]. \end{aligned} \tag{2.4}$$

Then $\{C(u)\}$ $(M_0 < n \leqq N_0)$ are mutually disjoint, and

$$\sum_{M_0 < u \leqq N_0} C(u) = \sum_{M_0 < u \leqq N_0} A_u(\varphi(u)). \tag{2.5}$$

For each u $(M_0 < u \leqq N_0)$ take an n $(n_0 \leqq n < n_0')$ such that $m_n < u \leqq m_{n+}$ and put

$$\Delta^-_{m_n,u} = E[t: f_u(t) - f_{m_n}(t) \leqq 0]. \tag{2.6}$$

It is to be noticed that $C(u)$ and $\Delta^-_{m_n,u}$ are not independent, but it can be shown by computations[7] that there exists a constant $c_{11} > 0$ such that

$$Pr(C(u) \cdot \Delta^-_{m_n,u}) > c_{11} Pr(C(u)). \tag{2.7}$$

---

[7] We sketch the proof of (2.7): Let us put

$$C(u, k) = E[t: f_u(t) = k; \quad f_v(t) \leqq \varphi(v), \quad u < v \leqq N_0],$$

where $k > \varphi(u)$ is an integer or integer $+ \frac{1}{2}$ according as u is even or odd. Then a simple calculation with binomial coefficients shows that

$$Pr\Big(\sum_{\varphi(u) < k \leqq \varphi(u) + u/\varphi(u)} C(u, k)\Big) > c_{46} Pr(C(u)).$$

Thus it suffices to show that, for $\varphi(u) < k \leqq \varphi(u) + u/\varphi(u)$,

$$Pr(C(u, k) \cdot \Delta^-_{m_n,u}) > c_{47} Pr(C(u, k)).$$

Now, it is easy to see that

$$\frac{Pr(C(u, k))}{Pr(C(u, k) \cdot \Delta^-_{m_n,u})} < c_{48} \binom{u}{\frac{u}{2} + k} \Big/ \binom{m_n}{\frac{u}{2} + k};$$

and a simple calculation shows that

$$\binom{u}{\frac{u}{2} + k} > c_{49} \binom{m_n}{\frac{u}{2} + k},$$

which completes the proof of (2.7).

On the other hand, since $t \in C(u)\cdot\Delta^{-}_{m_n,u}$ implies $f_{m_n}(t) \geqq f_u(t) > \varphi(u) \geqq \varphi(m_n)$, we have

$$C(u)\cdot\Delta^{-}_{m_n,u} \subset A_{m_n}(\varphi(m_n)) \tag{2.8}$$

for $m_n < u \leqq m_{n+1}$. Hence, since $\{C(u)\cdot\Delta^{-}_{m_n,u}\}$ $(M_0 < u \leqq N_0)$ are mutually disjoint, we have, by (2.7) and (2.5),

$$\begin{aligned}\sum_{M_0<m_n\leqq N_0} Pr(A_{m_n}(\varphi(m_n))) &\geqq \sum_{M_0<u\leqq N_0} Pr(C(u)\cdot\Delta^{-}_{m_n,u}) \\ &\geqq c_{11}\sum_{M_0<u\leqq N_0} Pr(C(u)) = c_{11}Pr\Big(\sum_{M_0<u\leqq N_0} C(u)\Big) \\ &= c_{11}Pr\Big(\sum_{M_0<u\leqq N_0} A_u(\varphi(u))\Big) > c_{10}\cdot c_{11} > 0.\end{aligned} \tag{2.9}$$

Since c_{10} and c_{11} are absolute positive constants, and since $M_0 = m_{n_0}$ can be taken arbitrarily large, this means that the set of all t for which the inequality (2.1) holds for infinitely many n, has positive measure. By the law of 0 or 1, this set must have measure 1, and thus Theorem 2 is proved.

3

THEOREM 3. *Let $\varphi(n)$ be monotone increasing and let us assume that*

$$\varphi(m_{n+1}) - \varphi(m_n) > c_{12}\,(m_n/\log\log m_n)^{\frac{1}{2}}. \tag{3.1}$$

Then a necessary and sufficient condition that $\varphi(n)$ belong to the lower class is that

$$\sum_{n=1}^{\infty} Pr(A_{m_n}(\varphi(m_n))) = \infty. \tag{3.2}$$

PROOF. The necessity follows from Theorem 1, without assuming (3.1). In order to prove that the condition (3.1) is sufficient, let us assume that $\varphi(n)$ is monotone increasing and satisfies (3.1) and (3.2). We first notice that (3.1) and (0.12) imply

$$\varphi(m_{n+1}) - \varphi(m_n) > c_{13}((m_{n+1}\log\log m_{n+1})^{\frac{1}{2}} - (m_n\log\log m_n)^{\frac{1}{2}}), \tag{3.3}$$

and hence

$$\varphi(m_n) > c_{14}(m_n\log\log m_n)^{\frac{1}{2}}. \tag{3.4}$$

From (3.4) and (0.7) it follows easily that

$$\lim_{n\to\infty} Pr(A_{m_n}(\varphi(m_n))) = 0. \tag{3.5}$$

Next we notice that we may assume

$$\varphi(n) \leqq (n\log\log n)^{\frac{1}{2}} \tag{3.6}$$

for sufficiently large n. Indeed, otherwise we may consider $\varphi_1(n) = \min(\varphi(n), (n\log\log n)^{\frac{1}{2}})$ instead of $\varphi(n)$. Since $\varphi_0(n) = (n\log\log n)^{\frac{1}{2}}$ clearly satisfies (3.1), $\varphi_1(n)$ satisfies it too. Further, it is obvious that (3.2) is satisfied by $\varphi_1(n)$ whenever it is satisfied by $\varphi(n)$. Moreover, since $\varphi_0(n)$ belongs to the

upper class, by the corollary to Theorem 1, $\varphi_1(n)$ belongs to the lower class at the same time as $\varphi(n)$.

Because of the law of 0 or 1, and because of Theorem 2, it is sufficient to prove that there exists a constant $c_{15} > 0$ such that there exists, for any $M_0 = m_{n_0}$, an $N_0 = m_{n_0'}$ $(n_0' > n_0)$ such that

$$Pr\Big(\sum_{M_0 < m_n \leqq N_0} A_{m_n}(\varphi(m_n))\Big) > c_{15}. \tag{3.7}$$

Let $\delta > 0$ be a small positive number, which we shall determine later. Then, by (3.2) and (3.4), there exists an N such that, for any $M_0 = m_{n_0} > N$, an $N_0 = m_{n_0'}$ $(n_0' > n_0)$ exists such that

$$\delta < \sum_{M_0 < m_n \leqq N_0} Pr(A_{m_n}(\varphi(m_n))) < 2\delta. \tag{3.8}$$

We shall prove that if δ is chosen sufficiently small (but fixed), then (3.7) is satisfied, with the same integers M_0 and N_0 as in (3.8), by a suitable positive constant $c_{15} > 0$.

In order to prove this, let us first put

$$\begin{aligned} D(m_n) &= A_{m_n}(\varphi(m_n)) - A_{m_n}(\varphi(m_n)) \cdot \sum_{m_n < m_{n+r} \leqq N_0} A_{m_{n+r}}(\varphi(m_{n+r})) \\ &= E[t : f_{m_n}(t) > \varphi(m_n); f_{m_{n+r}}(t) \leqq \varphi(m_{n+r}),\ m_n < m_{n+r} \leqq N_0]. \end{aligned} \tag{3.9}$$

Then $\{D(m_n)\}$ $(M_0 < m_n \leqq N_0)$ are mutually disjoint, and

$$\sum_{M_0 < m_n \leqq N_0} D(m_n) = \sum_{M_0 < m_n \leqq N_0} A_{m_n}(\varphi(m_n)). \tag{3.10}$$

Let us further put

$$\begin{aligned} D_1(m_n) &= A_{m_n}(\varphi(m_n)) - A_{m_n}\left(\varphi(m_n) + \frac{c_{12}}{2}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}}\right) \\ &= E\left[t : \varphi(m_n) < f_{m_n}(t) \leqq \varphi(m_n) + \frac{c_{12}}{2}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}}\right]. \end{aligned} \tag{3.11}$$

Then a simple computation will show that[8]

[8] We have clearly

$$\frac{Pr(D_1(m_n))}{Pr(A_{m_n}(\varphi(m_n)))} = \frac{\sum'\binom{m_n}{u}}{\sum_{u > \varphi(m_n)}\binom{m_n}{u}}$$

where the dash indicates that u runs only over the interval

$$\varphi(m_n), \qquad \varphi(m_n) + \frac{c_{12}}{2}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}}.$$

A simple calculation shows that

$$\sum'\binom{m_n}{u} \Big/ \sum_{u > \varphi(m_n)}\binom{m_n}{u} > c_{16}$$

which proves (3.12).

$$Pr(D_1(m_n)) > c_{16}Pr(A_{m_n}(\varphi(m_n))). \tag{3.12}$$

Let us put

$$D_2(m_n) = D_1(m_n)\cdot E[t: f_{m_{n+r-1}}(t) - f_{m_{n+r}}(t) \leqq 0, r = 1, 2, \cdots, h], \tag{3.13}$$

where h is a positive integer which we shall determine later. Then it is easy to see that

$$Pr(D_2(m_n)) \geqq 2^{-h}Pr(D_1(m_n)), \tag{3.14}$$

and that $t \in D_2(m_n)$ implies

$$f_{m_{n+r}}(t) \leqq f_{m_n}(t) < \varphi(m_{n+r}), \tag{3.15}$$

for $r = 1, 2, \cdots, h$. Let us further put

$$D_3(m_n) = D_2(m_n)\cdot E[t: f_{m_{n+r}}(t) \leqq \varphi(m_{n+r}),\ m_{n+h} < m_{n+r} \leqq N_0]. \tag{3.16}$$

Then it is clear that $D_3(m_n) \subset D(m_n) \subset A_{m_n}(\varphi(m_n))$. In order to complete the proof of Theorem 3, it is sufficient to prove that, if δ is chosen sufficiently small and if h is chosen sufficiently large (but both fixed), then there exists a constant $c_{17} > 0$ such that

$$Pr(D_3(m_n)) > c_{17}Pr(A_{m_n}(\varphi(m_n))). \tag{3.17}$$

Indeed, (3.17) will imply

$$\begin{aligned} Pr\Big(\sum_{M_0 < m_n \leqq N_0} A_{m_n}(\varphi(m_n))\Big) &= Pr\Big(\sum_{M_0 < m_n \leqq N_0} D(m_n)\Big) \\ &= \sum_{M_0 < m_n \leqq N_0} Pr(D(m_n)) \geqq \sum_{M_0 < m_n \leqq N_0} Pr(D_3(m_n)) \\ &> c_{17} \sum_{M_0 < m_n \leqq N_0} Pr(A_{m_n}(\varphi(m_n))) > c_{17}\cdot\delta, \end{aligned} \tag{3.18}$$

which means that (3.7) is satisfied by $c_{15} = c_{17}\cdot\delta > 0$, thus completing the proof of Theorem 3.

The rest of the proof of Theorem 3 is devoted to establishing the relation (3.17). For this purpose, put

$$D_{3,r}(m_n) = D_2(m_n)\cdot A_{m_{n+r}}(\varphi(m_{n+r})), \tag{3.19}$$

for all integers r such that $m_{n+h} < m_{n+r} \leqq N_0$. It is easy to see that

$$D_2(m_n) \subset D_3(m_n) + \sum_{m_{n+h} < m_{n+r} \leqq N_0} D_{3,r}(m_n). \tag{3.20}$$

We shall evaluate $Pr(\sum_{m_{n+h} < m_{n+r} \leqq N_0} D_{3,r}(m_n))$ by decomposing the sum into three parts: $\sum_{m_{n+h} < m_{n+r} \leqq 2m_n}$, $\sum_{2m_n < m_{n+r} \leqq m_n \log m_n}$, and $\sum_{m_n \log m_n < m_{n+r} \leqq N_0}$.

In the first place, $t \in D_{3,r}(m_n)$ implies

$$
\begin{aligned}
f_{m_{n+r}}(t) - f_{m_{n+h}}(t) &\geqq f_{m_{n+r}}(t) - f_{m_n}(t) \\
&> \varphi(m_{n+r}) - \varphi(m_n) - \frac{c_{12}}{2}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}} \\
&> \sum_{k=0}^{r-1} c_{12}\left(\frac{m_{n+k}}{\log\log m_{n+k}}\right)^{\frac{1}{2}} - \frac{c_{12}}{2}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}} \\
&> \frac{c_{12}\, r}{2}\left(\frac{m_n}{\log\log m_n}\right).
\end{aligned} \tag{3.21}
$$

Hence

$$
Pr(D_{3,r}(m_n)) \leqq \alpha_r \cdot Pr(D_2(m_n)), \tag{3.22}
$$

where

$$
\alpha_r = Pr\left(E\left[t : f_{m_{n+r}}(t) - f_{m_{n+h}}(t) > \frac{c_{12}\, r}{2}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}}\right]\right)
$$

$$
= Pr\left(A_{m_{n+r}-m_{n+h}}\left(\frac{c_{12}\, r}{2}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}}\right)\right) \tag{3.23}
$$

$$
< c_2 \frac{(m_{n+r} - m_{n+h})^{\frac{1}{2}}}{\frac{c_{12}\, r}{2}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}}} \exp\left[-\frac{2\left(\frac{c_{12}\, r}{2}\right)^2 \frac{m_n}{\log\log m_n}}{m_{n+r} - m_{n+h}}\right].
$$

Since, on the other hand, $m_{n+h} < m_{n+r} \leqq 2m_n$ implies

$$
\begin{aligned}
m_{n+r} - m_{n+h} &\leqq \sum_{k=0}^{r-1} (m_{n+k+1} - m_{n+k}) \\
&< c_5 r \frac{m_{n+r}}{\log\log m_{n+r}} \leqq 2c_5 r \frac{m_n}{\log\log m_n},
\end{aligned} \tag{3.24}
$$

we have, by (0.7),

$$
\alpha_r < c_{18} e^{-c_{19} r} \tag{3.25}
$$

for $m_{n+h} < m_{n+r} \leqq 2m_n$. Consequently,

$$
Pr\Big(\sum_{m_{n+h} < m_{n+r} \leqq 2m_n} D_{3,r}(m_n)\Big) < c_{18} \cdot Pr(D_2(m_n)) \cdot \sum_{r=h+1}^{\infty} e^{-c_{19} r}. \tag{3.26}
$$

Secondly, $t \in D_{3,r}(m_n)$ and $2m_n < m_{n+r} \leqq m_n \log m_n$ imply

$$
\begin{aligned}
f_{m_{n+r}}(t) - f_{m_{n+h}}(t) &> \frac{c_{12}}{2} \sum_{k=0}^{r-1} \left(\frac{m_{n+k}}{\log\log m_{n+k}}\right)^{\frac{1}{2}} \\
&> \frac{c_{12}}{2c_7} \sum_{k=0}^{r-1} \left((m_{n+k+1} \log\log m_{n+k+1})^{\frac{1}{2}} - (m_{n+k} \log\log m_{n+k})^{\frac{1}{2}}\right) \\
&> c_{20}(m_{n+r} \log\log m_{n+r})^{\frac{1}{2}}.
\end{aligned} \tag{3.27}
$$

Hence

$$Pr(D_{3,r}(m_n)) < \beta_r \cdot Pr(D_2(m_n)) \tag{3.28}$$

for $2m_n < m_{n+r} \leqq m_n \log m_n$, where

$$\begin{aligned}\beta_r &= Pr(E[t: f_{m_{n+r}}(t) - f_{m_{n+h}}(t) > c_{20}(m_{n+r} \log\log m_{n+r})^{\frac{1}{2}}]) \\ &= Pr(A_{m_{n+r}-m_{n+h}}(c_{20}(m_{n+r} \log\log m_{n+r})^{\frac{1}{2}})) \\ &< c_2 \frac{(m_{n+r} - m_{n+h})^{\frac{1}{2}}}{c_{20}(m_{n+r} \log\log m_{n+r})^{\frac{1}{2}}} \exp\left[-\frac{2c_{20}^2 m_{n+r} \log\log m_{n+r}}{m_{n+r} - m_{n+h}}\right] \qquad (3.29) \\ &< c_{21} e^{-c_{22} \log\log m_{n+r}} < \frac{c_{21}}{(\log m_n)^{c_{22}}}.\end{aligned}$$

On the other hand, the number of m_{n+r}'s satisfying $2m_n < m_{n+r} \leqq m_n \log m_n$ does not exceed $c_{23}(\log\log m_n)^2$.[9] Hence we have

$$Pr\Big(\sum_{2m_n < m_{n+r} \leqq m_n \log m_n} D_{3,r}(m_n)\Big) < Pr(D_2(m_n)) \cdot c_{24} \frac{(\log\log m_n)^2}{(\log m_n)^{c_{21}}}. \tag{3.30}$$

Lastly, $t \in D_{3,r}(m_n)$ and $m_n \log m_n < m_{n+r} \leqq N_0$ imply

$$\begin{aligned}f_{m_{n+r}}(t) - f_{m_{n+h}}(t) &> \varphi(m_{n+r}) - \varphi(m_n) - \frac{c_{12}}{2}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}} \qquad (3.31) \\ &> \varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}}.\end{aligned}$$

Hence

$$Pr(D_{3,r}(m_n)) < \gamma_r \cdot Pr(D_2(m_n)) \tag{3.32}$$

for $m_n \log m_n < m_{n+r} \leqq N_0$, where

$$\begin{aligned}\gamma_r &= Pr(E[t: f_{m_{n+r}}(t) - f_{m_{n+h}}(t) > \varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}}]) \\ &= Pr(A_{m_{n+r}-m_{n+h}}(\varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}})) \\ &< c_2 \frac{(m_{n+r} - m_{n+h})^{\frac{1}{2}}}{\varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}}} \\ &\qquad \cdot \exp\left[-\frac{2(\varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}})^2}{m_{n+r} - m_{n+h}}\right] \qquad (3.33) \\ &< c_2 \frac{(m_{n+r})^{\frac{1}{2}}}{\varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}}} \\ &\qquad \cdot \exp\left[-\frac{2(\varphi(m_{n+r}))^2}{m_{n+r}} + \frac{8\varphi(m_{n+r})(m_n \log\log m_n)^{\frac{1}{2}}}{m_{n+r}}\right].\end{aligned}$$

[9] It follows from (0.11) that the number of m_n's in the interval $(x, 2x)$ does not exceed $c_{50} \log\log x$. Thus the number of m_n's in the interval $(x, x \log x)$ does not exceed

$$c_{50} \log\log x \frac{\log\log x}{\log 2} < c_{13}(\log\log x)^2.$$

On the other hand, for sufficiently large n, $m_{n+r} > m_n \log m_n$ implies

$$\varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}} > \tfrac{1}{2}\varphi(m_{n+r}), \tag{3.34}$$

$$\varphi(m_{n+r})(m_n \log\log m_n)^{\frac{1}{2}} < m_{n+r}\,. \tag{3.35}$$

Hence

$$\begin{aligned}\gamma_r &< c_{25}\cdot\frac{(m_{n+r})^{\frac{1}{2}}}{\varphi(m_{n+r})}\exp\left[-\frac{2(\varphi(m_{n+r}))^2}{m_{n+r}}\right] \\ &< c_{26}\,Pr(A_{m_{n+r}}(\varphi(m_{n+r})))\,.\end{aligned} \tag{3.36}$$

Consequently,

$$\begin{aligned}&Pr\Big(\sum_{m_n\log m_n<m_{n+r}\leqq N_0} D_{3,r}(m_n)\Big) \\ &\qquad< Pr(D_2(m_n))\cdot c_{25}\sum_{M_0<m_{n+r}\leqq N_0} Pr(A_{m_{n+r}}(\varphi(m_{n+r}))) \\ &\qquad< c_{26}\cdot 2\delta\cdot Pr(D_2(m_n))\,.\end{aligned} \tag{3.37}$$

Combining (3.26), (3.30) and (3.37), we have finally

$$\begin{aligned}&Pr\Big(\sum_{m_{n+h}<m_{n+r}\leqq N_0} D_{3,r}(m_n)\Big) \\ &\qquad< Pr(D_2(m_n))\left\{c_{18}\cdot\sum_{r=h+1}^{\infty} e^{-c_{19}r} + c_{24}\frac{(\log\log m_n)^2}{(\log m_n)^{c_{22}}} + c_{26}\cdot 2\delta\right\}.\end{aligned} \tag{3.38}$$

Hence, if we take h sufficiently large and δ sufficiently small, then we have

$$Pr\Big(\sum_{m_{n+h}<m_{n+r}\leqq N_0} D_{3,r}(m_n)\Big) < \theta\cdot Pr(D_2(m_n)), \tag{3.39}$$

where θ is a constant with $0 < \theta < 1$. Consequently, by (3.20),

$$\begin{aligned}&Pr(D_3(m_n)) > (1-\theta)\cdot Pr(D_2(m_n)) \\ &\qquad> 2^{-h}(1-\theta)\cdot Pr(D_1(m_n)) > (1-\theta)\cdot c_{27}Pr(A_{m_n}(\varphi(m_n)) \text{ (by 3.12)} \\ &\qquad> c_{17}Pr(A_{m_n}(\varphi(m_n)))\,,\end{aligned} \tag{3.40}$$

which proves (3.17). The proof of Theorem 3 is completed.

COROLLARY 1. *$\varphi(n) = (1/\sqrt{2}+\epsilon)(n\log\log n)^{\frac{1}{2}}$ belongs to the lower class for $\epsilon \leqq 0$.*

COROLLARY 2. *The expression* (0.6) *belongs to the lower class for $\epsilon \leqq 0$.*

PROOF. Follows immediately from Theorem 3 and (0.7).

4

THEOREM 4. *Let $\varphi(n)/n^{\frac{1}{2}}$ be monotone increasing. Then a necessary and sufficient condition that $\varphi(n)$ belong to the lower class is that*

$$\sum_{n=1}^{\infty} Pr(A_{m_n}(\varphi(m_n))) = \infty\,. \tag{4.1}$$

We need the following

Hence

$$Pr(D_{3,r}(m_n)) < \beta_r \cdot Pr(D_2(m_n)) \tag{3.28}$$

for $2m_n < m_{n+r} \leqq m_n \log m_n$, where

$$\begin{aligned}\beta_r &= Pr(E[t: f_{m_{n+r}}(t) - f_{m_{n+h}}(t) > c_{20}(m_{n+r} \log\log m_{n+r})^{\frac{1}{2}}]) \\ &= Pr(A_{m_{n+r}-m_{n+h}}(c_{20}(m_{n+r} \log\log m_{n+r})^{\frac{1}{2}})) \\ &< c_2 \frac{(m_{n+r} - m_{n+h})^{\frac{1}{2}}}{c_{20}(m_{n+r} \log\log m_{n+r})^{\frac{1}{2}}} \exp\left[-\frac{2c_{20}^2 m_{n+r} \log\log m_{n+r}}{m_{n+r} - m_{n+h}}\right] \qquad (3.29) \\ &< c_{21} e^{-c_{22} \log\log m_{n+r}} < \frac{c_{21}}{(\log m_n)^{c_{22}}}.\end{aligned}$$

On the other hand, the number of m_{n+r}'s satisfying $2m_n < m_{n+r} \leqq m_n \log m_n$ does not exceed $c_{23}(\log\log m_n)^2$.[9] Hence we have

$$Pr\Big(\sum_{2m_n < m_{n+r} \leqq m_n \log m_n} D_{3,r}(m_n)\Big) < Pr(D_2(m_n)) \cdot c_{24} \frac{(\log\log m_n)^2}{(\log m_n)^{c_{21}}}. \tag{3.30}$$

Lastly, $t \in D_{3,r}(m_n)$ and $m_n \log m_n < m_{n+r} \leqq N_0$ imply

$$\begin{aligned}f_{m_{n+r}}(t) - f_{m_{n+h}}(t) &> \varphi(m_{n+r}) - \varphi(m_n) - \frac{c_{12}}{2}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}} \qquad (3.31) \\ &> \varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}}.\end{aligned}$$

Hence

$$Pr(D_{3,r}(m_n)) < \gamma_r \cdot Pr(D_2(m_n)) \tag{3.32}$$

for $m_n \log m_n < m_{n+r} \leqq N_0$, where

$$\begin{aligned}\gamma_r &= Pr(E[t: f_{m_{n+r}}(t) - f_{m_{n+h}}(t) > \varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}}]) \\ &= Pr(A_{m_{n+r}-m_{n+h}}(\varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}})) \\ &< c_2 \frac{(m_{n+r} - m_{n+h})^{\frac{1}{2}}}{\varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}}} \\ &\qquad \cdot \exp\left[-\frac{2(\varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}})^2}{m_{n+r} - m_{n+h}}\right] \qquad (3.33) \\ &< c_2 \frac{(m_{n+r})^{\frac{1}{2}}}{\varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}}} \\ &\qquad \cdot \exp\left[-\frac{2(\varphi(m_{m+r}))^2}{m_{n+r}} + \frac{8\varphi(m_{n+r})(m_n \log\log m_n)^{\frac{1}{2}}}{m_{n+r}}\right].\end{aligned}$$

[9] It follows from (0.11) that the number of m_n's in the interval $(x, 2x)$ does not exceed $c_{50} \log\log x$. Thus the number of m_n's in the interval $(x, x \log x)$ does not exceed

$$c_{50} \log\log x \frac{\log\log x}{\log 2} < c_{23}(\log\log x)^2.$$

On the other hand, for sufficiently large n, $m_{n+r} > m_n \log m_n$ implies

$$\varphi(m_{n+r}) - 2(m_n \log\log m_n)^{\frac{1}{2}} > \tfrac{1}{2}\varphi(m_{n+r}), \tag{3.34}$$

$$\varphi(m_{n+r})(m_n \log\log m_n)^{\frac{1}{2}} < m_{n+r}\,. \tag{3.35}$$

Hence

$$\begin{aligned}\gamma_r &< c_{25}\cdot\frac{(m_{n+r})^{\frac{1}{2}}}{\varphi(m_{n+r})}\exp\left[-\frac{2(\varphi(m_{n+r}))^2}{m_{n+r}}\right] \\ &< c_{26}\,Pr(A_{m_{n+r}}(\varphi(m_{n+r})))\,.\end{aligned} \tag{3.36}$$

Consequently,

$$\begin{aligned}&Pr\Big(\sum_{m_n\log m_n<m_{n+r}\leqq N_0} D_{3,r}(m_n)\Big) \\ &\qquad < Pr(D_2(m_n))\cdot c_{25}\sum_{M_0<m_{n+r}\leqq N_0} Pr(A_{m_{n+r}}(\varphi(m_{n+r}))) \\ &\qquad < c_{26}\cdot 2\delta\cdot Pr(D_2(m_n))\,.\end{aligned} \tag{3.37}$$

Combining (3.26), (3.30) and (3.37), we have finally

$$\begin{aligned}&Pr\Big(\sum_{m_{n+h}<m_{n+r}\leqq N_0} D_{3,r}(m_n)\Big) \\ &\qquad < Pr(D_2(m_n))\left\{c_{18}\cdot\sum_{r=h+1}^{\infty} e^{-c_{19}r} + c_{24}\frac{(\log\log m_n)^2}{(\log m_n)^{c_{22}}} + c_{26}\cdot 2\delta\right\}.\end{aligned} \tag{3.38}$$

Hence, if we take h sufficiently large and δ sufficiently small, then we have

$$Pr\Big(\sum_{m_{n+h}<m_{n+r}\leqq N_0} D_{3,r}(m_n)\Big) < \theta\cdot Pr(D_2(m_n)), \tag{3.39}$$

where θ is a constant with $0 < \theta < 1$. Consequently, by (3.20),

$$\begin{aligned}Pr(D_3(m_n)) &> (1-\theta)\cdot Pr(D_2(m_n)) \\ &> 2^{-h}(1-\theta)\cdot Pr(D_1(m_n)) > (1-\theta)\cdot c_{27}Pr(A_{m_n}(\varphi(m_n)) \text{ (by 3.12)} \\ &> c_{17}Pr(A_{m_n}(\varphi(m_n)))\,,\end{aligned} \tag{3.40}$$

which proves (3.17). The proof of Theorem 3 is completed.

COROLLARY 1. *$\varphi(n) = (1/\sqrt{2}+\epsilon)(n\log\log n)^{\frac{1}{2}}$ belongs to the lower class for $\epsilon \leqq 0$.*

COROLLARY 2. *The expression* (0.6) *belongs to the lower class for $\epsilon \leqq 0$.*

PROOF. Follows immediately from Theorem 3 and (0.7).

4

THEOREM 4. *Let $\varphi(n)/n^{\frac{1}{2}}$ be monotone increasing. Then a necessary and sufficient condition that $\varphi(n)$ belong to the lower class is that*

$$\sum_{n=1}^{\infty} Pr(A_{m_n}(\varphi(m_n))) = \infty\,. \tag{4.1}$$

We need the following

LEMMA 1. *Let $M_1 < N_1 < M_2 < N_2 < \cdots < M_i < N_i < \cdots$ be a sequence of positive integers tending to infinity, and let $\varphi(n)$ be such that*

$$\varphi(m_{n+1}) - \varphi(m_n) > c_{28}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}} \qquad M_i < m_n \qquad m_{n+1} \leqq N_i, \tag{4.2}$$

$$\varphi(m_n) > c_{29}(m_n \log\log m_n)^{\frac{1}{2}}, \qquad M_i < m_n < m_{n+1} \leqq N_i, \tag{4.3}$$

$$\sum_{M_i < m_n \leqq N_i} Pr(A_{m_n}(\varphi(m_n))) > c_{30}. \tag{4.4}$$

Then $\varphi(n)$ belongs to the lower class.

We do not give the proof of Lemma 1, since it can be carried out in the same way as in Theorem 3.

PROOF OF THEOREM 4. The necessity of the condition is clear by Theorem 1. In order to prove that it is sufficient, let us assume that $\varphi(n)/n^{\frac{1}{2}}$ is monotone increasing and that (4.1) is satisfied. We shall prove that there exists a sequence of integers $M_1 < N_1 < M_2 < N_2 \quad \cdots < M_i < N_i < \cdots$ tending to infinity, which satisfies the conditions of Lemma 1.

If we have

$$\varphi(m_n) < \tfrac{1}{10}(m_n \log\log m_n)^{\frac{1}{2}} \tag{4.5}$$

for all sufficiently large n, then the fact that $\varphi_0(n) = \frac{1}{10}(n \log\log n)^{\frac{1}{2}}$ belongs to the lower class (see Corollary 1 to Theorem 3), together with Theorem 2, will imply that $\varphi(n)$ belongs to the lower class. On the other hand, if

$$\varphi(m_n) > \tfrac{1}{20}(m_n \log\log m_n)^{\frac{1}{2}} \tag{4.6}$$

for sufficiently large n, then

$$\varphi(m_{n+1}) \geqq \left(\frac{m_{n+1}}{m_n}\right)^{\frac{1}{2}} \varphi(m_n) > \left(1 + \frac{c_{31}}{\log\log m_n}\right)\varphi(m_n) \tag{4.7}$$

by (0.11), and hence

$$\varphi(m_{n+1}) - \varphi(m_n) > c_{31}\frac{\varphi(m_n)}{\log\log m_n} > \frac{c_{31}}{20}\left(\frac{m_n}{\log\log m_n}\right)^{\frac{1}{2}}. \tag{4.8}$$

Consequently, by Theorem 3, $\varphi(n)$ must belong to the lower class again.

Thus, in order to prove Theorem 4, we have only to consider the case when there exist two sequences of integers tending to infinity $\{M_i\} = \{m_{n_i}\}$ $(i = 1, 2, \cdots)$ and $\{N_i\} = \{m_{n_i'}\}$ $(i = 1, 2, \cdots)$ such that $M_1 < N_1 < M_2 < N_2 < \cdots < M_i < N_i < \cdots$, and

$$\varphi(M_i) = \varphi(m_{n_i}) \geqq \tfrac{1}{10}(M_i \log\log M_i)^{\frac{1}{2}}, \tag{4.9}$$

$$\varphi(N_i) = \varphi(m_{n_i}) \leqq \tfrac{1}{20}(N_i \log\log N_i)^{\frac{1}{2}}. \tag{4.10}$$

We may assume that

$$\varphi(m_n) < \tfrac{1}{10}(m_n \log\log m_n)^{\frac{1}{2}} \tag{4.11}$$

for $M_i < m_n \leqq N_i$ (i.e. for $n_i < n \leqq n_i'$) $(i = 1, 2, \cdots)$.

We shall prove that the conditions of Lemma 1 are all satisfied by these $\{M_i\}$ $(i = 1, 2, \cdots)$ and $\{N_i\}$ $(i = 1, 2, \cdots)$. Since $\varphi(M_i)/(M_i)^{\frac{1}{4}} \leqq \varphi(N_i)/N_i^{\frac{1}{4}}$ by assumption, we have $\frac{1}{10}(\log\log M_i)^{\frac{1}{4}} \leqq \frac{1}{20}(\log\log N_i)^{\frac{1}{4}}$ for $i = 1, 2, \cdots$. Since $M_i, N_i \to \infty$ as $i \to \infty$, it follows that we have $M_i^4 \leqq N_i$ for sufficiently large i.

Let now $M_i < m_n < N_i$. Then

$$
\begin{aligned}
Pr(A_{m_n}(\varphi(m_n))) &> c_1 \frac{(m_n)^{\frac{1}{4}}}{\varphi(m_n)} e^{-2(\varphi(m_n))^2/m_n} \\
&> c_1 \frac{10}{(\log\log m_n)^{\frac{1}{4}}} e^{-(\log\log m_n)/50} = \frac{10\, c_1}{(\log\log m_n)^{\frac{1}{4}}(\log m_n)^{1/50}} \\
&> \frac{1}{(\log m_n)^{1/49}}
\end{aligned} \tag{4.12}
$$

for sufficiently large i. Since $2\cdot\log M_i\cdot\log\log M_i < n < 3\cdot\log M_i\cdot\log\log M_i$ implies $\log M_i < n/\log n < 4\log M_i$, or equivalently $M_i < e^{n/\log n} < M_i^4$, for sufficiently large i, we have

$$
\begin{aligned}
\sum_{M_i < m_n \leqq N_i} Pr(A_{m_n}(\varphi(m_n))) &> \sum_{M_i < m_n \leqq N_i} \frac{1}{(\log m_n)^{1/49}} \\
&> \sum_{M_i < m_n \leqq M_i^4} \frac{1}{(\log m_n)^{1/49}} > \sum_{2p_i < n \leqq 3p_i} \frac{1}{n^{1/49}},
\end{aligned} \tag{4.13}
$$

where $p_i = \log N_i\cdot\log\log N_i$. Thus (4.4) is satisfied. (4.3) is clearly satisfied with $c_{29} = \frac{1}{20}$; (4.8) shows that (4.2) is also satisfied. This completes the proof of Theorem 4.

5

THEOREM 5. *Let $\varphi(n)$ satisfy*

$$\varphi(n) > c_{32}(n\log\log n)^{\frac{1}{4}}, \tag{5.1}$$

$$\sum_{n=1}^{\infty} Pr(A_{m_n}(\varphi(m_n))) = \infty. \tag{5.2}$$

Then $\varphi(n)$ belongs to the lower class.

To prove Theorem 5 we need the following

LEMMA 2. *Let $\varphi(n)$ be monotone increasing, and let $\{m_{n_i}\}$ $(i = 1, 2, \cdots)$ be a subsequence of $\{m_n\}$ $(n = 1, 2, \cdots)$ such that*

$$\varphi(m_{n_{i+1}}) \geqq \varphi(m_{n_i}) + c_{33}((m_{n_{i+1}}\log\log m_{n_{i+1}})^{\frac{1}{4}} - (m_{n_i}\log\log m_{n_i})^{\frac{1}{4}}) \tag{5.3}$$

$$\sum_{i=1}^{\infty} Pr(A_{m_{n_i}}(\varphi(m_{n_i}))) = \infty. \tag{5.4}$$

Then $\varphi(n)$ belongs to the lower class.

Since the proof of Lemma 2 can be carried out exactly as in the proof of Theorem 3, we omit the proof.

PROOF OF THEOREM 5. As in the proof of Theorem 3, we may assume that

$$\varphi(n) \leqq (n \log\log n)^{\frac{1}{2}} \tag{5.5}$$

for sufficiently large n. We shall find a subsequence $\{m_{n_i}\}$ $(i = 1, 2, \cdots)$ of $\{m_n\}$ $(n = 1, 2, \cdots)$ which satisfies the conditions of Lemma 2. For this purpose we classify the integers m_n into two classes. The first class I consists of all integers m_p for which

$$\varphi(m_q) \geqq \varphi(m_p) + \epsilon((m_q \log\log m_q)^{\frac{1}{2}} - (m_p \log\log m_p)^{\frac{1}{2}}) \tag{5.6}$$

for all $q \geqq p$, where ϵ is a positive constant with $0 < \epsilon < c_{32}$ which we shall determine later. All other integers m_p will belong to the second class II. We shall prove that, if we denote by $\{m_{n_i}\}$ $(i = 1, 2, \cdots, m_{n_i} < m_{n_{i+1}})$ the integers of the class I, then this sequence satisfies the conditions of Lemma 2. Indeed, (5.3) is clear from (5.6). In order to prove (5.4) for the m_{n_i}'s of the class I, let us denote by II_i the set of all integers m_p of the class II such that $m_p < m_{n_i}$ and

$$\varphi(m_{n_i}) < \varphi(m_p) + \epsilon((m_{n_i} \log\log m_{n_i})^{\frac{1}{2}} - (m_p \log\log m_p)^{\frac{1}{2}}). \tag{5.7}$$

By definition, for each m_p of the class II, there exists an m_q $(m_q > m_p)$ such that

$$\varphi(m_q) < \varphi(m_p) + \epsilon((m_q \log\log m_q)^{\frac{1}{2}} - (m_p \log\log m_p)^{\frac{1}{2}}). \tag{5.8}$$

Because of (5.1) and the relation $\epsilon < c_{32}$, there exists, for each m_p of II, a largest integer m_q $(m_q > m_p)$ satisfying (5.8). This m_q clearly belongs to I. Hence we have $\sum_{i=1}^{\infty} \mathrm{II}_i = \mathrm{II}$ (II_i are not necessarily mutually disjoint).

Thus in order to prove (5.4), we need only prove that there exists a constant $c_{34} > 0$ such that

$$\sum_{m_p \in \mathrm{II}_i} Pr(A_{m_p}(\varphi(m_p))) < c_{34} Pr(A_{m_{n_i}}(\varphi(m_{n_i}))). \tag{5.9}$$

For this purpose we shall first show that

$$m_{n_i} < c_{35} m_p \tag{5.10}$$

for all $m_p \in \mathit{II}_i$, where c_{35} is independent of i and p. Indeed, if (5.10) is false, we have

$$\begin{aligned} \varphi(m_{n_i}) &< \varphi(m_p) + \epsilon((m_{n_i} \log\log m_{n_i})^{\frac{1}{2}} - (m_p \log\log m_p)^{\frac{1}{2}}) \\ &< (m_p \log\log m_p)^{\frac{1}{2}} + \epsilon(m_{n_i} \log\log m_{n_i})^{\frac{1}{2}} \\ &< \left(\frac{1}{\sqrt{c_{35}}} + \epsilon\right)(m_{n_i} \log\log m_{n_i})^{\frac{1}{2}}, \end{aligned} \tag{5.11}$$

and this is a contradiction to (5.1) if c_{35} is sufficiently large.

By (5.5), (5.7) and (5.10), if $m_p = m_{n_i-k} \in \mathrm{II}_i$, then we have

$$Pr(A_{m_p}(\varphi(m_p))) < c_2 \frac{m_p^{\frac{1}{2}}}{\varphi(m_p)} e^{-2(\varphi(m_p))^2/m_p}$$

$$< \frac{c_2}{(\log\log m_p)^{\frac{1}{2}}} \cdot \exp\left[-\frac{2\{\varphi(m_{n_i}) - \epsilon((m_{n_i}\log\log m_{n_i})^{\frac{1}{2}} - (m_p \log\log m_p)^{\frac{1}{2}})\}^2}{m_{n_i}}\right]$$

$$(5.12) \quad < \frac{c_2}{(\log\log m_p)^{\frac{1}{2}}} \cdot \exp\left[-\frac{2(\varphi(m_{n_i}))^2 - 4\epsilon\varphi(m_{n_i})((m_{n_i}\log\log m_{n_i})^{\frac{1}{2}} - (m_p\log\log m_p)^{\frac{1}{2}})}{m_{n_i} - k\cdot c_5 \dfrac{m_p}{\log\log m_p}}\right]$$

$$< \frac{c_{36}}{(\log\log m_n)^{\frac{1}{2}}} \exp\left[-\frac{2(\varphi(m_{n_i}))^2}{m_{n_i}}\right]\cdot\eta^2 < c_{37}\cdot Pr(A_{m_{n_i}}(\varphi(m_{n_i})))\cdot\eta^2$$

where

$$\eta = \exp\left[\frac{(\varphi(m_{n_i}))^2}{m_{n_i}} - \frac{(\varphi(m_{n_i}))^2 - 2\epsilon\varphi(m_{n_i})((m_{n_i}\log\log m_{n_i})^{\frac{1}{2}} - (m_p\log\log m_p)^{\frac{1}{2}})}{m_{n_i} - kc_{38}\dfrac{m_{n_i}}{\log\log m_{n_i}}}\right]$$

$$< \exp\left[\frac{(\varphi(m_{n_i}))^2}{m_{n_i}} - \frac{(\varphi(m_{n_i}))^2 - 2\epsilon\varphi(m_{n_i})((m_{n_i}\log\log m_{n_i})^{\frac{1}{2}} - (m_p\log\log m_p)^{\frac{1}{2}})}{m_{n_i}} \cdot\left(1 + k\frac{c_{39}}{\log\log m_{n_i}}\right)\right]$$

(5.13)

$$< \exp\left[\frac{2\epsilon\varphi(m_{n_i})}{m_{n_i}}((m_{n_i}\log\log m_{n_i})^{\frac{1}{2}} - (m_p\log\log m_p)^{\frac{1}{2}}) - \frac{kc_{40}(\varphi(m_{n_i}))^2}{m_{n_i}\log\log m_{n_i}}\right]$$

$$< \exp\left[\frac{2\epsilon\varphi(m_{n_i})}{m_{n_i}}\cdot k\cdot c_7\left(\frac{m_{n_i}}{\log\log m_{n_i}}\right)^{\frac{1}{2}} - k\frac{c_{40}(\varphi(m_{n_i}))^2}{m_{n_i}\log\log\ m_{n_i}}\right]$$

$$< \exp\left[(2\epsilon\cdot c_7 c_{32} - c_{40}\cdot c_{32}^2)\cdot k\right].$$

Hence, if we take ϵ sufficiently small, then

$$\eta < e^{-c_{42}k} \tag{5.14}$$

with a positive constant c_{42}. Hence

$$\begin{aligned}\sum_{m_p \in \Pi_i} Pr(A_{m_p}(\varphi(m_p))) &< c_{37}\, Pr(A_{m_{n_i}}(\varphi(m_{n_i}))) \cdot \sum_{k=1}^{\infty} e^{-2c_{42}k} \\ &= \frac{c_{37}}{1 - e^{-2c_{42}}}\, Pr(A_{m_{n_i}}(\varphi(m_{n_i})))\end{aligned} \tag{5.15}$$

which proves (5.9). This completes the proof of Theorem 5.

Before concluding this chapter, let us add some more results without proof.

1). *If $\varphi(n)$ is monotone increasing and belongs to the lower class, then $\varphi(n) + c\,(n/\log\log n)^{\frac{1}{2}}$ belongs to the lower class for all c.*

This result is the best possible. For, if $\psi(n) \to \infty$, then we can find a monotone function $\varphi(n)$ belonging to the lower class such that $\varphi(n) + \psi(n)(n/\log\log n)^{\frac{1}{2}}$ belongs to the upper class.

2). *If $\varphi(n)$ is monotone increasing and belongs to the lower class, then $\varphi(n) + c(n/\varphi(n))$ belongs to the lower class for all c.* Since we can always assume that $\varphi(n) < (n \log\log n)^{\frac{1}{2}}$, 2) is slightly stronger than 1).

3). *Let $\varphi(n)$ be monotone increasing, and suppose that it belongs to the upper class. Then for almost all t, there exist only finitely many n such that for some $m < n$, $|f_n(t) - f_m(t)| > \varphi(n)$.*

4). *For almost all t, we have*

$$\limsup_{n\to\infty} \frac{\sum_{k=1}^{n} f_k(t)}{\frac{1}{2} n^{\frac{1}{2}} \left(\frac{\log\log n}{2}\right)^{\frac{1}{2}}} = 1. \tag{5.16}$$

Professor J. L. Doob suggested that if $n_1 < n_2 < \cdots$ is a sequence of integers with $n_{i+1}/n_i > c > 1$, then for almost all t,

$$\lim_{p\to\infty} \frac{1}{p} \sum_{i=1}^{p} \frac{\sum_{k=1}^{n_i} f_k(t)}{n_i^{\frac{1}{2}}} = 0. \tag{5.17}$$

Indeed, it is not difficult to show that (5.17) holds. In fact, the condition $n_{i+1}/n_i > c > 1$ can be weakened, but it is necessary that n_i tends to infinity with a certain speed (quicker than i).

5). *There exists a continuous strictly decreasing function $\psi(x)$ defined for $0 \leqq x \leqq 1$, with $\psi(0) = 1$, $\psi(1) = 0$, such that, for almost all t, the upper density of the set of n's for which*

$$f_n(t) \geqq x\left(\frac{n \log\log n}{2}\right)^{\frac{1}{2}} \tag{5.18}$$

is exactly $\psi(x)$.[10]

[10] This problem was suggested by W. Ambrose.

6

In this final chapter, we shall construct an increasing function $\varphi(n)$ such that

$$\sum_{n=1}^{\infty} Pr(A_{m_n}(\varphi(m_n))) = \infty, \tag{6.1}$$

and nevertheless $\varphi(n)$ belongs to the upper class. This shows that the *converse of Theorem* 1 *is not true.*

We put

$$p_k = 2^{2^{2^k}}, \qquad k = 1, 2, \cdots, \tag{6.2}$$

and define $\varphi(n)$ as follows:

$$\varphi(n) = \log k \cdot \sqrt{p_k}, \qquad \text{for } p_{k-1} < n \leqq p_k. \tag{6.3}$$

It follows from (0.11) that the number of m_n's satisfying $\frac{1}{2}p_k < m_n \leqq p_k$ is $\geqq c_{43} \log\log p_k$ and hence $\geqq c_{44}2^k$. Consequently, from (0.7) we have

$$\sum_{\frac{1}{2}p_k < m_n \leqq p_k} Pr(A_{m_n}(\varphi(m_n))) > \frac{c_1}{\log k} e^{-4(\log k)^2} \cdot c_{44} 2^k \geqq c_{45} > 0. \tag{6.4}$$

Since this is true for each k, (6.1) is proved.

Denote now

$$M_k = E[t: \max_{1 \leqq n \leqq p_k} f_n(t) > \log k \cdot \sqrt{p_k}]. \tag{6.5}$$

In order to show that $\varphi(n)$ belongs to the upper class, it is clearly sufficient to prove that

$$\sum_{k=1}^{\infty} Pr(M_k) < \infty. \tag{6.6}$$

It is easy to see that[11]

$$Pr(M_k) \leqq 2Pr(E[t: f_{p_k}(t) > \log k \sqrt{p_k}]. \tag{6.7}$$

---

[11] In general, we have

$$Pr(E[t: \max_{1 \leqq n \leqq p} f_n(t) > x]) \leqq 2\, Pr(E[t: f_p(t) > x]).$$

Indeed, we have

$$\begin{aligned} Pr(E[t: \max_{1 \leqq n \leqq p} f_n(t) > x]) &= Pr(E[t: f_p(t) > x]) \\ &\quad + \sum_{n=1}^{p-1} Pr(E[t: f_1(t) \leqq x, \cdots, f_{n-1}(t) \leqq x, f_n(t) > x, f_p(t) \leqq x]) \\ &= Pr(E[t: f_p(t) > x]) \\ &\quad + \sum_{n=1}^{p-1} Pr(E[t: f_1(t) \leqq x, \cdots, f_{n-1}(t) \leqq x, f_n(t) > x, f_p(t) \geqq 2f_n(t) - x]) \\ &\leqq Pr(E[t: f_p(t) > x]) \\ &\quad + \sum_{n=1}^{p-1} Pr(E[t: f_1(t) \leqq x, \cdots, f_{n-1}(t) \leqq x, f_n(t) > x, f_p(t) > x]) \\ &= 2\, Pr(E[t: f_p(t) > x]). \end{aligned}$$

Thus from (0.7) we have

$$\sum_{k=1}^{\infty} Pr(M_k) \leqq 2 \cdot c_2 \cdot \sum_{k=1}^{\infty} \frac{1}{\log k} e^{-2(\log k)^2} < \infty, \tag{6.8}$$

which proves (6.6).

My indebtedness to my friend S. Kakutani is very great. In fact, he wrote the whole paper after listening to my rough oral exposition.

UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PA.

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

ON AN ELEMENTARY PROOF OF SOME ASYMPTOTIC FORMULAS IN THE THEORY OF PARTITIONS

BY P. ERDÖS

(Received January 7, 1942)

Denote by $p(n)$ the number of partitions of n. Hardy and Ramanujan[1] proved in their classical paper that

$$p(n) \sim \frac{1}{4n3^{\frac{1}{2}}}\, e^{cn^{\frac{1}{2}}}, \qquad c = \pi(\tfrac{2}{3})^{\frac{1}{2}},$$

using complex function theory. The main purpose of the present paper is to give an elementary proof of this formula. But we can only prove with our elementary method that

$$p(n) \sim \frac{a}{n}\, e^{cn^{\frac{1}{2}}} \tag{1}$$

and are unable to prove that $a = 1/4.3^{\frac{1}{2}}$.

Our method will be very similar to that used in a previous paper.[2] The starting point will be the following identity:

$$np(n) = \sum_{v=1}\sum_{k=1} vp(n - kv), \qquad p(0) = p(-m) = 0. \tag{2}$$

(We easily obtain (2) by adding up all the $p(n)$ partitions of n, and noting that v occurs in $p(n - v)$ partitions.) (2) is of course well known. In fact, Hardy and Ramanujan state in their paper[3] that by using (2) they have obtained an elementary proof of

$$\log p(n) \sim cn^{\frac{1}{2}}. \tag{3}$$

The proof of (3) is indeed easy. First we show that

$$p(n) < e^{cn^{\frac{1}{2}}}. \tag{4}$$

We use induction. (4) clearly holds for $n = 1$. By (2) and the induction hypothesis we have

$$np(n) < \sum_{\substack{v=1\\ kv<n}}\sum_{k=1} ve^{c(n-kv)^{\frac{1}{2}}} < \sum_{v=1}^{\infty}\sum_{k=1}^{\infty} ve^{cn^{\frac{1}{2}}-ckv/2n^{\frac{1}{2}}} = e^{cn^{\frac{1}{2}}} \sum_{k=1}^{\infty} \frac{e^{-kc/2n^{\frac{1}{2}}}}{(1 - e^{kc/2n^{\frac{1}{2}}})^2}.$$

[1] Hardy, Ramanujan, *Asymptotic formulae in combinatory analysis*, Proc. London Math. Soc. 17, (1918), pp. 75–115.

[2] Erdös, *On some asymptotic formulas in the theory of factorisatio numerorum*, these Annals 42, (1941), pp. 989–993.

[3] Hardy, Ramanujan, ibid, p. 79.

Now it is easy to see that for all real x, $\frac{e^{-x}}{(1-e^{-x})^2} < \frac{1}{x^2}$. Thus

$$np(n) < e^{cn^{\frac{1}{2}}} \sum_{k=1}^{\infty} \frac{4n}{c^2 k^2} = ne^{cn^{\frac{1}{2}}},$$

which proves (4).

Similarly but with slightly longer calculations, we can prove that for every $\epsilon > 0$ there exists an $A > 0$ such that

$$p(n) > \frac{1}{A} e^{(c-\epsilon)n^{\frac{1}{2}}}. \tag{5}$$

(4) and (5) clearly imply (3).

To prove (1) we need the following

LEMMA 1:

$$\sum = \sum_{\substack{v=1 \\ kv<n}}^{\infty} \sum_{k=1}^{\infty} \frac{ve^{c(n-kv)^{\frac{1}{2}}}}{n-kv} = e^{cn^{\frac{1}{2}}} \left[1 + O\left(\frac{1}{n^{\frac{1}{2}+\epsilon}}\right)\right], \tag{6}$$

for some fixed $\epsilon > 0$.

PROOF. We omit as many details as possible, since the proof is quite straight forward and uninteresting. We evidently have by expanding $1/(n-kv)$ and omitting the terms with $kv > n^{\frac{1}{2}+\epsilon}$

$$\sum_{\substack{v=1 \\ kv<n}}^{n} \sum_{k=1}^{n} \frac{ve^{c(n-kv)^{\frac{1}{2}}}}{n-kv} = \frac{1}{n} \sum_{\substack{v=1 \\ kv<n}}^{\infty} \sum_{k=1}^{\infty} ve^{c(n-kv)^{\frac{1}{2}}} + \frac{1}{n^2} \sum_{\substack{v=1 \\ kv<n}}^{\infty} \sum_{k=1}^{\infty} kv^2 e^{c(n-kv)^{\frac{1}{2}}}$$

$$+ O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right) = \sum_1 + \sum_2 + O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right).$$

Now

$$\sum_2 = \frac{e^{cn^{\frac{1}{2}}}}{n^2} \sum_{v=1}^{\infty} \sum_{k=1}^{\infty} kv^2 e^{-ckv/2n^{\frac{1}{2}}} + O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right).$$

(It is easy to see that the other terms of $e^{c(n-kv)^{\frac{1}{2}}}$ can be neglected and that the summation for v and k can be extended to ∞.) Thus

$$\sum_2 = \frac{e^{cn^{\frac{1}{2}}}}{n^2} \sum_{k=1}^{\infty} \frac{2k}{(1-e^{-kc/2n^{\frac{1}{2}}})^3} + O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right) = e^{cn^{\frac{1}{2}}} \sum_{k=1}^{\infty} \frac{2k \cdot 8n^{\frac{3}{2}}}{k^3 c^3}$$

$$+ O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right) = \frac{4}{c} \frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}}} + O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right).$$

On the other hand

$$\Sigma_1 = \frac{e^{cn^{\frac{1}{2}}}}{n} \sum_{v=1}^{\infty} \sum_{k=1}^{\infty} v e^{-ckv/2n^{\frac{1}{2}} - ck^2v^2/8n^{\frac{1}{2}}} + O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right)$$

$$= \frac{e^{cn^{\frac{1}{2}}}}{n} \left(\sum_{v=1}^{\infty} \sum_{k=1}^{\infty} v e^{-ckv/2n^{\frac{1}{2}}} - \frac{ck^2v^3}{8n^{\frac{1}{2}}} e^{-ckv/2n^{\frac{1}{2}}} \right) = \Sigma_1' - \Sigma_1'' + O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right).$$

$$\Sigma_1'' = \frac{ce^{cn^{\frac{1}{2}}}}{8n^{\frac{1}{2}}} \sum_{v=1}^{\infty} \sum_{k=1}^{\infty} k^2 v^3 e^{-ckv/2n^{\frac{1}{2}}} = \frac{ce^{cn^{\frac{1}{2}}}}{8n^{\frac{1}{2}}} \sum_{k=1}^{\infty} \frac{6k^2}{(1-e^{-ck/2n^{\frac{1}{2}}})^4} + O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right)$$

$$= \frac{ce^{cn^{\frac{1}{2}}}}{8n^{\frac{1}{2}}} \sum_{k=1}^{\infty} \frac{6k^2 \cdot 16n^2}{k^4c^4} + O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right) = \frac{3}{c} \frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}}} + O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right).$$

$$\Sigma_1' = \frac{e^{cn^{\frac{1}{2}}}}{n} \sum_{v=1}^{\infty} \sum_{k=1}^{\infty} v e^{-ckv/2n^{\frac{1}{2}}} = \frac{e^{cn^{\frac{1}{2}}}}{n} \sum_{k=1}^{\infty} \frac{e^{-ck/2n^{\frac{1}{2}}}}{(1 - e^{-ck/2n^{\frac{1}{2}}})^2}.$$

A simple calculation shows that

$$\frac{e^{-x}}{(1 - e^{-x})^2} = \frac{1}{x^2} + O(1), \quad \text{i.e.} \quad \frac{e^{-ck/2n^{\frac{1}{2}}}}{(1 - e^{-ck/2n^{\frac{1}{2}}})^2} = \frac{4n}{c^2k^2} + O(1).$$

Hence

$$\Sigma_1' = \frac{e^{cn^{\frac{1}{2}}}}{n} \sum_{k=1}^{u} \frac{4n}{c^2k^2} + \sum_{k>u} \frac{e^{-ck/2n^{\frac{1}{2}}}}{(1 - e^{-ck/2n^{\frac{1}{2}}})^2} + O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right), \qquad u = [n^{\frac{1}{2}}].$$

But

$$\sum_{k=1}^{u} \frac{4n}{c^2k^2} = \frac{4n}{c^2} \frac{\pi^2}{6} - \frac{4n}{c^2} \sum_{k>u} \frac{1}{k^2} = n - \frac{4n}{c^2u} + O\left(\frac{n}{u^2}\right).$$

And

$$\sum_{k>u} \frac{e^{-ck/2n^{\frac{1}{2}}}}{(1 - e^{-ck/2n^{\frac{1}{2}}})^2} = \int_u^{\infty} \frac{e^{-cx/2n^{\frac{1}{2}}}}{(1 - e^{-cx/2n^{\frac{1}{2}}})^2} dx + O\left(\frac{1}{u^2}\right)$$

$$= \frac{2n^{\frac{1}{2}}}{c(1 - e^{-cu/2n^{\frac{1}{2}}})} - \frac{n^{\frac{1}{2}}}{c} + O\left(\frac{1}{u^2}\right) = \frac{4n}{c^2u} - \frac{n^{\frac{1}{2}}}{c} + O\left(\frac{1}{n^{\frac{1}{2}+\epsilon}}\right).$$

Thus finally

$$\Sigma_1' = e^{cn^{\frac{1}{2}}} - \frac{e^{cn^{\frac{1}{2}}}}{cn^{\frac{1}{2}}} + O\left(\frac{e^{cn^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}}\right).$$

Hence

$$\Sigma = \Sigma_1' - \Sigma_1'' + \Sigma_2 = e^{cn^{\frac{1}{2}}}\left[1 + O\left(\frac{1}{n^{\frac{1}{2}+\epsilon}}\right)\right]$$

which proves the lemma.

Next we show that

$$0 < \liminf \frac{np(n)}{e^{cn^{\frac{1}{2}}}} \leqq \limsup \frac{np(n)}{e^{cn^{\frac{1}{2}}}} < \infty . \tag{7}$$

To prove (7) write

$$c_1^{(n)} = \max_{m \leqq n} \frac{mp(m)}{e^{cm^{\frac{1}{2}}}} . \tag{8}$$

Clearly by (8) and (6) and (2)

$$(n+1)p(n+1) \leqq c_1^{(n)} \sum_{\substack{v=1 \\ kv<n}} \sum_{k=1} \frac{ve^{c(n+1-kv)^{\frac{1}{2}}}}{n+1-kv} < c_1^{(n)} e^{c(n+1)^{\frac{1}{2}}} \left(1 + \frac{b_1}{n^{\frac{1}{2}+\epsilon}}\right).^4$$

Write

$$\frac{(n+j)p(n+j)}{e^{c(n+j)^{\frac{1}{2}}}} = c_1^{(n)} \left(1 + \frac{b_j}{n^{\frac{1}{2}+\epsilon}}\right), \qquad j = 1, 2, \cdots .$$

Then

$$\begin{aligned}(n+r+1)p(n+r+1) &< c_1^{(n)} \sum_{\substack{v=1 \\ kv \leqq n+r}} \sum_{k=1} \frac{ve^{c(n+r+1-kv)^{\frac{1}{2}}}}{n+r+1-kv} \\ &\quad + c_1^{(n)} \frac{\max_{j \leqq r} b_j}{n^{\frac{1}{2}+\epsilon}} \sum_{\substack{v=1 \\ kv \leqq r}} \sum_{k=1} \frac{ve^{c(n+r+1-kv)^{\frac{1}{2}}}}{n+r+1-kv} \\ &< c_1^{(n)} e^{c(n+r+1)^{\frac{1}{2}}} \left(1 + \frac{b_1}{n^{\frac{1}{2}+\epsilon}} + \frac{\max_{j \leqq r} b_j}{n^{\frac{1}{2}+\epsilon}} \frac{r^2 e^{c(n+r+1)^{\frac{1}{2}}}}{n}\right),\end{aligned}$$

since

$$\sum_{kv \leqq r} v \leqq r^2 .$$

Hence

$$b_{r+1} < b_1 + \frac{r^2 \max_{j \leqq r} b_j}{n} .$$

We show that, for $r^2 \leqq n/2$, $b_{r+1} \leqq 2b_1$. We use induction. We have

$$b_{r+1} < b_1 + \frac{r^2 \cdot 2b_1}{n} \leqq 2b_1 .$$

---

[4] b_1 is chosen such that for every $m > 0$

$$\sum_{\substack{v \\ m-kv>0}} \sum_{k} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m-kv} < e^{c(m^{\frac{1}{2}})} \left(1 + \frac{b_1}{m^{\frac{1}{2}+\epsilon}}\right).$$

Thus

$$c_1^{[n+(\frac{1}{2}n)^{\frac{1}{2}}]} \leqq c_1^{(n)}\left(1 + \frac{2b_1}{n^{\frac{1}{2}+\epsilon}}\right).$$

Or

$$c_1^{((m+1)^2)} < c_1^{(m^2)}\left(1 + \frac{10b_1}{n^{\frac{1}{2}+\epsilon}}\right);$$

and since $\sum m^{1/1+\epsilon}$ converges we see that $\limsup c_1^{(n)} < \infty$; i.e. $\limsup np(n)/e^{cn^{\frac{1}{2}}} < \infty$. Similarly we can show that $\liminf np(n)/e^{cn^{\frac{1}{2}}} > 0$, which completes the proof of (7).

Next we prove that

$$\liminf \frac{np(n)}{e^{cn^{\frac{1}{2}}}} = \limsup \frac{np(n)}{e^{cn^{\frac{1}{2}}}} \tag{9}$$

and this will complete the proof of (1).

Suppose that (9) does not hold; write

$$\liminf \frac{np(n)}{e^{cn^{\frac{1}{2}}}} = d, \qquad \limsup \frac{np(n)}{e^{cn^{\frac{1}{2}}}} = D. \tag{10}$$

Now choose n large and such that

$$\frac{np(n)}{e^{cn^{\frac{1}{2}}}} > D - \epsilon.$$

Then since $p(n)$ is an increasing function of n there exists a c_2 such that for every m in the range $n \leqq m \leqq n + c_2 n^{\frac{1}{2}}$

$$\frac{mp(m)}{e^{cm^{\frac{1}{2}}}} > \frac{d + D}{2}.$$

Now we claim that for every r_1 there exists a $\delta_{r_1} = \delta(r_1)$ such that, for $n \leqq m \leqq n + r_1 n^{\frac{1}{2}}$,

$$\frac{mp(m)}{e^{cm^{\frac{1}{2}}}} > d + \delta_{r_1}. \tag{11}$$

We prove (11) as follows: We evidently have by our lemma

$$mp(m) \geqq d \sum_{\substack{v=1 \\ kv<m}} \sum_{k=1} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m - kv} + \frac{D - d}{2} \sum_{n \leqq m-kv \leqq n+c_2 n^{\frac{1}{2}}} \sum_{v=1} \sum_{k=1} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m - kv} - o(e^{cm^{\frac{1}{2}}})$$ [5]

[5] The term $o(e^{cm^{\frac{1}{2}}})$ is present because d is the lower limit and not the lower bound of $\frac{mp(m)}{e^{cm^{\frac{1}{2}}}}$.

$$> de^{cm^{\frac{1}{2}}} + \frac{D-d}{2}\frac{e^{cn^{\frac{1}{2}}}}{m} \sum_{n \leqq m-v \leqq n+c_2 n^{\frac{1}{2}}} v - o(e^{cm^{\frac{1}{2}}}) > de^{cm^{\frac{1}{2}}} + c_3 e^{cn^{\frac{1}{2}}} - o(e^{cm^{\frac{1}{2}}})$$

$$> (d + \delta_{r_1})e^{cm^{\frac{1}{2}}}, \qquad \left(\text{i.e. } \frac{e^{cn^{\frac{1}{2}}}}{e^{cm^{\frac{1}{2}}}} > c_4\right).$$

which proves (11).

Suppose $2n \geqq m \geqq n + sn^{\frac{1}{2}}$, s sufficiently large; we show that

$$\sum_{\substack{v=1 \\ m-kv<n}} \sum_{k=1} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m-kv} < \frac{e^{cm^{\frac{1}{2}}}}{s^{10}}. \tag{12}$$

Clearly

$$\sum_{0<m-kv<n} \sum_{k} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m-kv} \leqq \sum_{kv>sn^{\frac{1}{2}}} \sum_{k} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m-kv}$$

$$< e^{cm^{\frac{1}{2}}} \sum_{\frac{1}{2}m>kv>sn^{\frac{1}{2}}} \sum_{k} \frac{2ve^{-ckv/2m^{\frac{1}{2}}}}{m} + \sum_{m>kv\geqq\frac{1}{2}m} \sum_{k} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m-kv}$$

$$< e^{cm^{\frac{1}{2}}} \sum_{\frac{1}{2}m>kv>sn^{\frac{1}{2}}} \sum_{k} \frac{2ve^{-ckv/2m^{\frac{1}{2}}}}{m} + m^2 e^{c(\frac{1}{2}m)^{\frac{1}{2}}},$$

since

$$\sum_{kv<x} \sum_{k} v \leqq x^2.$$

Further

$$\sum_{\frac{1}{2}m>kv>sn^{\frac{1}{2}}} \sum_{k} ve^{-ckv/2m^{\frac{1}{2}}} < \sum_{u=1}^{m} \sum_{(u+1)sn^{\frac{1}{2}}\geqq kv>usn^{\frac{1}{2}}} \sum_{v} \sum_{k} ve^{-cusn^{\frac{1}{2}}/2m^{\frac{1}{2}}}$$

$$< \sum_{u=1}^{m} \sum_{kv\leqq(u+1)sn^{\frac{1}{2}}} \sum_{v=1} \sum_{k=1} ve^{-cus/4} < \sum_{u=1}^{m} (u+1)^2 s^2 n e^{-cus/4}.$$

Thus

$$\sum_{\frac{1}{2}m>kv>sn^{\frac{1}{2}}} \sum_{k} ve^{-ckv/2m^{\frac{1}{2}}} < ms^2 \sum_{u=1}^{\infty} (u+1)^2 e^{-cus/4} < \frac{m}{4s^{10}}$$

for sufficiently large s. Hence finally

$$\sum_{\substack{v=1 \\ m-kv<n}} \sum_{k=1} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m-kv} < \frac{e^{cm^{\frac{1}{2}}}}{2s^{10}} + m^2 e^{c(\frac{1}{2}m)^{\frac{1}{2}}} < \frac{e^{cm^{\frac{1}{2}}}}{s^{10}}$$

for sufficiently large m and s (since $s < n^{\frac{1}{2}}$).

Consider now the intervals $n + tn^{\frac{1}{4}}$, $n + (t+1)n^{\frac{1}{4}}$, $t > r_1$, $t + 1 < n^{\frac{1}{4}}$. Split it into t^2 equal parts. Write

$$\min \frac{mp(m)}{e^{cm^{\frac{1}{4}}}} = d + \delta_t^u, \qquad n \leqq m \leqq n + \left(t + \frac{u+1}{t^2}\right) n^{\frac{1}{4}}$$

and put $\delta_t^{t^2-1} = \delta_t$. Now let $n + (t + u/t^2)n^{\frac{1}{4}} \leqq m \leqq n + (t + (u+1)/t^2)n^{\frac{1}{4}}$; then we have

$$mp(m) > d \sum_{\substack{v=1 \\ kv<m}} \sum_{k=1} \frac{ve^{c(m-kv)^{\frac{1}{4}}}}{m - kv} + \delta_t^{(u-1)} \sum_v{}' \sum_k{}' \frac{ve^{c(m-kv)^{\frac{1}{4}}}}{m - kv} - o(e^{cm^{\frac{1}{4}}}),$$

where the primes indicate that the summation is extended only over those v and k for which $n \leqq m - kv \leqq n + (t + u/t^2)n^{\frac{1}{4}}$. Further by Lemma 1

$$mp(m) \geqq (d + \delta_t^{(u-1)})e^{cm^{\frac{1}{4}}} - \delta_t^{(u-1)} \sum{}'' \frac{ve^{c(m-kv)^{\frac{1}{4}}}}{m - kv}$$

$$- \delta_t^{(u-1)} \sum{}''' \frac{ve^{c(m-kv)^{\frac{1}{4}}}}{m - kv} - o(e^{cm^{\frac{1}{4}}}),$$

where in $\sum''$ the summation is extended only over those v and k for which $m - kv \leqq n$, and in $\sum'''$ the summation is extended only over those v and k for which $m - kv \geqq n + (t + u/t^2)n^{\frac{1}{4}}$. We have by (11)

$$\sum'' < \frac{e^{cm^{\frac{1}{4}}}}{t^{10}}.$$

Further we have

$$\sum''' < \frac{n}{t^4} \frac{2e^{cm^{\frac{1}{4}}}}{m} < \frac{2e^{cm^{\frac{1}{4}}}}{t^4}.$$

Hence finally

$$mp(m) > e^{cm^{\frac{1}{4}}} \left(d + \delta_t^{(u-1)} - \frac{3\delta_t^{(u-1)}}{t^4}\right) - o(e^{cm^{\frac{1}{4}}}).$$

Hence

$$\delta_t^{(u)} > \delta_t^{(u-1)} \left(1 - \frac{3}{t^4}\right) - o(1).$$

Thus if t is fixed, independent of n, we have

$$\delta_{t+1} > \delta_t \left(1 - \frac{3}{t^4}\right)^{t^2} - o(1),$$

therefore

$$\delta_t > \delta_{r_1} \prod_{u > r_1} \left(1 - \frac{3}{u^4}\right)^{u^2} - o(1).$$

But $\prod_u (1 - 3/u^4)^{u^2}$ converges; thus, if r_1 was sufficiently large, we have $\delta_t > \delta_{r_1}/2$. Now choose r_2 sufficiently large; then we have $\delta_{r_2} > \delta_{r_1}/2$, i.e. for $n \leqq m \leqq n + r_2 n^{\frac{1}{2}}$,

$$\frac{mp(m)}{e^{cm^{\frac{1}{2}}}} > d + \frac{\delta_{r_1}}{2}.$$

Consider the interval $n + tn^{\frac{1}{2}}, n + (t+1)n^{\frac{1}{2}}, t > r_2$. Split it into t^2 equal parts. $\delta_t^{(u)}$ and δ_t have the same meaning as before. Suppose $n + (t + u/t^2)n^{\frac{1}{2}} \leqq m \leqq n + (t + (u+1)/t^2)n^{\frac{1}{2}}$; then evidently

$$mp(m) > (d + \delta_t^{(u-1)}) \sum_v{}' \sum_k{}' \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m - kv},$$

where the primes indicate that the summation is extended only over those v and k for which $n \leqq m - kv \leqq n + n^{\frac{1}{2}}(t + u/t^2)$.
Now

$$\sum_v{}' \sum_k{}' \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m - kv} = \sum_{\substack{v=1 \\ kv<m}} \sum_{k=1} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m - kv} - \sum'' - \sum''',$$

where $\sum''$ and $\sum'''$ are defined as before. By (12) and the previous estimate of $\sum'''$ we have

$$\sum'' < \frac{e^{cm^{\frac{1}{2}}}}{t^{10}}, \qquad \sum''' < \frac{2e^{cm^{\frac{1}{2}}}}{t^4}.$$

Hence by Lemma 1

$$mp(m) > e^{cm^{\frac{1}{2}}}(d + \delta_t^{(u-1)})\left(1 - \frac{3}{t^4}\right) - \frac{b_1(d + \delta_t^{(u-1)})e^{cm^{\frac{1}{2}}}}{n^{\frac{1}{2}+\epsilon}};$$

i.e.

$$d + \delta_t^{(u)} > (d + \delta_t^{(u-1)})\left(1 - \frac{3}{t^4}\right) - \frac{b_1(d + \delta_t^{(u-1)})}{n^{\frac{1}{2}+\epsilon}},$$

and

$$d + \delta_{t+1} > (d + \delta_t)\left(1 - \frac{3}{t^4}\right)^{t^2} - \frac{b_1 t^2(d + \delta_t^{(u-1)})}{n^{\frac{1}{2}+\epsilon}},$$

or

$$d + \delta_s > \left(d + \frac{\delta_{r_1}}{2}\right) \prod_{t>r_2} \left(1 - \frac{3}{t^4}\right)^{t^2} - \frac{b_2 s^3}{n^{\frac{1}{2}+\epsilon}}.$$

For sufficiently large r_2 we have,

$$\left(d + \frac{\delta_{r_1}}{2}\right) \prod_{t>r_2} \left(1 - \frac{3}{t^4}\right)^{t^2} > d + \frac{\delta_{r_1}}{4},$$

and if $s \leqq (\log n)^2$ and n is sufficiently large,

$$\delta_s > \frac{\delta_{r_1}}{8};$$

that is, for $n \leqq m \leqq n + n^{\frac{1}{2}}(\log n)^2$

$$\frac{mp(m)}{e^{cm^{\frac{1}{2}}}} > d + \frac{\delta_{r_1}}{8}.$$

Now suppose $m > n + n^{\frac{1}{2}}(\log n)^2$; we shall show that

$$\sum = \sum_{\substack{v \\ 0<m-kv<n}} \sum_{k} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m - kv} < \frac{e^{cm^{\frac{1}{2}}}}{m}.$$

We have

$$\sum < m^2 e^{cn^{\frac{1}{2}}} < m^2 e^{cm^{\frac{1}{2}} - 10c \log m} < \frac{e^{cm^{\frac{1}{2}}}}{m}$$

for sufficiently large n. Hence by Lemma 1,

$$\sum_{\substack{v \\ m-kv \geqq n}} \sum_{k} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m - kv} > e^{cm^{\frac{1}{2}}}\left(1 - \frac{b_1'}{n^{\frac{1}{2}+\epsilon}}\right).$$ [6]

Now we continue as in the proof of (7). Suppose $t > n + n^{\frac{1}{2}}(\log n)^2$; write

$$d + \delta_t = \min \frac{mp(m)}{e^{cm^{\frac{1}{2}}}}, \qquad n \leqq m \leqq t.$$

Then

$$(t+1)p(t+1) \geqq (d + \delta_t) \sum_{\substack{v \\ t-kv \geqq n}} \sum_{k} \frac{ve^{c(t-kv)^{\frac{1}{2}}}}{t - kv} > (d + \delta_t)e^{ct^{\frac{1}{2}}}\left(1 - \frac{b_1'}{t^{\frac{1}{2}+\epsilon}}\right).$$

Write

$$\frac{(t+r)p(t+r)}{e^{c(t+r)^{\frac{1}{2}}}} = (d + \delta_t)\left(1 - \frac{b_r'}{t^{\frac{1}{2}+\epsilon}}\right).$$

Then as in the proof of (7) we have

$$(t+j+1)p(t+j+1) > (d + \delta_t) \sum_{\substack{v \\ t+j+1-kv \geqq n}} \sum_{k} \frac{ve^{c(t+j+1-kv)^{\frac{1}{2}}}}{t + j + 1 - kv}$$

$$- (d + \delta_t) \frac{\max_{r \leqq j} b_r'}{t^{\frac{1}{2}+\epsilon}} \frac{j^2}{t} e^{c(t+j+1)^{\frac{1}{2}}}$$

$$> (d + \delta_t)e^{c(t+j+1)^{\frac{1}{2}}}\left(1 - \frac{b_1'}{(t+j+1)^{\frac{1}{2}}}\right) - (d + \delta_t) \frac{\max_{r \leqq j} b_r'}{t^{\frac{1}{2}+\epsilon}} \frac{j^2}{t} e^{c(t+j+1)^{\frac{1}{2}}}$$

$$= (d + \delta_t)e^{c(t+j+1)^{\frac{1}{2}}}\left(1 - \frac{b_{j+1}'}{t^{\frac{1}{2}+\epsilon}}\right),$$

[6] As in footnote 4 b_1' is chosen such that for every $m > n + n^{\frac{1}{2}}(\log n)^2$

$$\sum_{\substack{v \\ n<m-kv}} \sum_{k} \frac{ve^{c(m-kv)^{\frac{1}{2}}}}{m - kv} > e^{cm^{\frac{1}{2}}}\left(1 - \frac{b_1'}{m^{\frac{1}{2}+\epsilon}}\right).$$

where

$$b'_{j+1} < b'_1 + \max_{r \leqq j} b'_r \cdot \frac{j^2}{t}.$$

We show that for $j^2 < t/2$ we have, $b'_{j+1} < 2b'_1$. We use induction; we have

$$b'_{j+1} < b'_1 + \frac{2b'_1}{2} = 2b'_1.$$

Thus

$$d + \delta_{[t+t^{\frac{1}{2}},t^{\frac{1}{2}}]} > (d + \delta_t)\left(1 - \frac{2b'_1}{t^{1+\epsilon}}\right).$$

That is,

$$d + \delta_{(s+1)^2} > (d + \delta_{s^2})\left(1 - \frac{10b'_1}{s^{1+\epsilon}}\right).$$

Therefore

$$d + \delta_{u^2} > \left(d + \frac{\delta_{r_1}}{8}\right) \prod_{v > \log n} \left(1 - \frac{10b'_1}{v^{1+\epsilon}}\right) > d + \frac{\delta_{r_1}}{10},$$

which contradicts (10); and this completes the proof of (1).

As can be seen, the main idea of our proof is rather simple; unfortunately the details are long and cumbersome. By the same method we can prove the following result: Let m be a fixed integer. Denote by $p^{(m)}_{a_1,a_2,\cdots,a_r}(n)$ the number of partitions of n into integers congruent to one of the numbers $a_1, a_2, \cdots a_r \pmod m$. Then

$$p^{(m)}_{a_1,a_2,\cdots,a_r}(n) \sim \frac{a}{n^\alpha} e^{cn^{\frac{1}{2}}}, \qquad ((a_1, a_2, \cdots, a_r^m) = 1) \tag{13}$$

where C depends on m and r, and α and a depend on $m, a_1, a_2, \cdots a_r$.

The same method will work if we consider partitions of n into rth powers. Denote the number of partitions of n into rth powers by $p_r(n)$, Hardy, Ramanujan and Wright[7] proved that

$$p_r^{(n)} \sim c_1 n^{1/(r+1)-\frac{3}{2}} e^{c_2 n^{1/(r+1)}}. \tag{14}$$

Clearly as in the case of $p(n)$ we have

$$np_r(n) = \sum_{\substack{v \\ vk<n}} \sum_{k} v^r p_r(n - kv^r).$$

[7] Hardy, Ramanujan, ibid. p. 111. Maitland Wright, Acta Math. 63, (1934), pp. 143-191. Wright proves a very much sharper result than (13).

To prove (14) we should only have to prove the analogue of our lemma, namely

$$\sum_{\substack{v \\ v^r k<n}} \sum_{k} (n - v^r k)^{1/(r+1)-\frac{1}{2}} e^{c_2(n-v^r k)^{1/(r+1)}}$$

(15)

$$= n^{1/(r+1)-\frac{1}{2}} e^{c_2 n^{1/(r+1)}} \left[1 + O\left(\frac{1}{n^{1-(1/(r+1))+\epsilon}}\right)\right].$$

If (15) is proved the proof of (14) proceeds as in the case of $p(n)$.

I have not worked out a proof of (15); it seems likely that a proof would be longer than that of Lemma 1, but would not present any particular difficulties.

Recently Ingham[8] proved a Tauberian theorem from which (1) and (14) follow as corollaries. In fact his Theorem 2 gives a more general result, from which (13) also follows as a very special case.

Denote by $P_r(n)$ the number of partitions of n into powers of r. Clearly

$$nP_r(n) = \sum_{\substack{v \\ r^v k<n}} \sum_{k} r^v P_r(n - r^v k).$$

It might be possible to get an asymptotic formula for $P_r(n)$ by our method. I have not succeeded so far. But we can show without difficulty that

$$\log P_r(n) \sim \frac{(\log n)^2}{2 \log r}. \tag{16}$$

We have

$$f(x) = \sum_{n=0}^{\infty} P_r(n) x^n = \prod_{v=1}^{\infty} \frac{1}{1 - x^{r^v}}.$$

It is easy to see that for $0 \leqq x \leqq 1$,

$$c_1 \left(\frac{1}{1-x}\right)^{(1/(2 \log a)) \log 1/(1-x)} < f(x) < c_2 \left(\frac{1}{1-x}\right)^{(1/(2 \log a)) \log 1/(1-x)}. \tag{17}$$

Thus

$$P_r(n) \left(1 - \frac{1}{n}\right)^n < f\left(1 - \frac{1}{n}\right) < c_2 n^{(\log n)/(2 \log a)};$$

that is

$$P_r(n) < c_3 n^{(\log n)/(2 \log a)}, \qquad \log P_r(n) < (1 - \epsilon) \frac{(\log n)^2}{2 \log a} \quad \text{for} \quad n > n_0.$$

Suppose now that for a certain large n $\log (P_r(n)) < (1 - \epsilon)(\log n)^2/2 \log a$; then, since for $m < n$ $P_r(m) \leqq P_r(n)$ we have

$$f(x) < e^{(1-\epsilon) \cdot (\log n)^2/(2 \log a)} \sum_{k=0}^{n} x^k + \sum_{k>n} c_3 k^{(\log k)/(2 \log a)} x^k, \tag{18}$$

---

[8] A. E. Ingham, *A Tauberian Theorem for Partitions*, these Annals, 42 (1941), p. 1083.

and a simple calculation shows that (18) contradicts (17). (Choose $x = (1 - \delta)n$, $\delta = \delta(\epsilon)$). The same method would of course give

$$\log (p(n)) \sim \pi \left(\frac{2n}{3}\right)^{\frac{1}{2}}.$$

We can also prove the following results:

I. Let $a_1 < a_2 < \cdots$ be an infinite sequence of integers of density α, such that the a's have no common factor. Denote by $p'(n)$ the number of partitions of n into the a's. Then

$$\log (p'(n)) \sim c(\alpha n)^{\frac{1}{2}}. \qquad (c = \pi(\tfrac{2}{3})^{\frac{1}{2}} \tag{19}$$

II. Let $a_1 < a_2 < \cdots$ be an infinite sequence of integers of density α, such that every sufficiently large m can be expressed as the sum of different a's. Then denote by $P'(n)$ the number of partitions of n into different a's. Then

$$\log P'(n) \sim c\left(\frac{\alpha}{2} n\right)^{\frac{1}{2}}. \tag{20}$$

We shall sketch the proof of II; the proof of I is similar but simpler. Denote by $P(n)$ the number of partitions of n into different summands: it is well known that[9]

$$\log P(n) \sim c\left(\frac{n}{2}\right)^{\frac{1}{2}}. \tag{21}$$

First we show that

$$\limsup \frac{\log P'(n)}{c\left(\frac{\alpha}{2} n\right)^{\frac{1}{2}}} \leqq 1. \tag{22}$$

To the partition $n = a_{i_1} + a_{i_2} + \cdots + a_{i_r}$ we make correspond the partition $i_1 + i_2 + \cdots + i_r$. For $i > i_0$ we have $i < a_i(\alpha + \epsilon)$ therefore $i_1 + i_2 + \cdots + i_k < n(\alpha + \epsilon) + i_0^2$. Thus each partition of n into the a's is mapped into a partition of integers $\leqq n(\alpha + 2\epsilon)$ with all integers as summands; hence from (20) we obtain (22). Next we prove that

$$\liminf \frac{\log P'(n)}{c\left(\frac{\alpha}{2} n\right)^{\frac{1}{2}}} \geqq 1. \tag{23}$$

Split the sequence a_i into two disjoint sequences $b_1, b_2, \cdots$ and $c_1, c_2, \cdots$ where the b's have density 0 and every sufficiently large integer is the sum of different b's and the c's are the remaining a's. It is easy to see that we can find the b's with the required property; also the density of the c's is clearly α. Denote by $Q(n)$ the number of partitions of n into the c's. Now associate

[9] A well known result of Euler states that the number of partitions of n into odd integers equals the number of partitions of n into different summands. Thus (20) follows from i.

with the partition $n = i_1 + i_2 + \cdots + i_k$, $i_1 < i_2 < \cdots < i_k$ the partition $c_{i_1} + c_{i_2} + \cdots + c_{i_k}$; as before, we have

$$\frac{n}{\alpha + \epsilon} < c_{i_1} + c_{i_2} + \cdots + c_{i_k} < \frac{n}{\alpha - \epsilon}.$$

Hence for at least one $n/(\alpha + \epsilon) < m < n/(\alpha - \epsilon)$, $Q(m) > p(n)(\alpha - \epsilon)/n$. Thus there exists a sequence of integers $x_1 < x_2 < \cdots$ with $\lim x_{i+1}/x_i = 1$ and

$$24) \qquad \liminf \frac{\log Q(x_i)}{c\left(\frac{\alpha}{2} x_i\right)^{\frac{1}{2}}} = 1.$$

Now suppose $x_j \leqq m < x_{j+1}$. Choose x_i such that $\epsilon m > m - x_i > C$. Then $m - x_i$ is a sum of different b's, hence $P(m) \geqq Q(x_i)$. Thus (23) follows from (24), and this completes the proof of II.

If might be worth while to mention the following problem: Let $a_1 < a_2 < \cdots$ be an infinite sequence of integers, such that $\log P(n) \sim c(\alpha n)^{\frac{1}{2}}$. Does it then follow that the density of the a's is α. I cannot decide this problem. Perhaps the following result might be of some interest in this connection: Let $a_1 < a_2 \cdots$ be an infinite sequence of integers. $f(n)$ denotes the number of a's $\leqq n$, and $\varphi(n)$ denotes the number of solutions of $a_i + a_j \leqq n$. It can be shown trivially that if $\lim f(n)/n^{\alpha} = c_1$ then $\lim \varphi(n)/n^{2\alpha} = c_2$. But the converse is also true, and can be simply proved by using a Tauberian theorem of Hardy and Littlewood.[10] We have

$$(f(z))^2 = \left(\sum_{i=1}^{\infty} z^{a_i}\right)^2 = \sum_{k=1}^{\infty} d_k z^k$$

and, since $\sum d_k = \varphi(n) \sim c_2 n^{2\alpha}$, we evidently have

$$f(z) \underset{L \to 1}{\sim} \frac{c_3}{(1 - z)^{\alpha}}$$

and hence by the theorem of Hardy and Littlewood,

$$f(n) = \sum_{a_k \leqq n} 1 \sim c_1 n^{\alpha}.$$

By the same methods that were used in proving II, we can prove the following result: Denote by $R(n)$ the number of partitions of n into integers relatively prime to n. We have

$$\log R(n) \sim c(\varphi(n))^{\frac{1}{2}}.$$

Similarly, if we denote by $R'(n)$ the number of partitions of n into different integers relatively prime to n, we have

$$\log R'(n) \sim c\left(\frac{\varphi(n)}{2}\right)^{\frac{1}{2}}.$$

---

[10] Hardy-Littlewood, *Tauberian Theorems*, Proc. London Math. Soc. 13, (1914), pp. 174–191.

I have not succeeded in finding asymptotic formulas for $R(n)$ and $R'(n)$. This problem seems rather difficult.

March 12, 1942.

In the meantime I have proved the above conjecture. Consider

$$f(x) = \sum_{n=1}^{\infty} P(n)x^n = \prod_{k=1}^{\infty} \frac{1}{1 - x^{a_k}}.$$

If we assume that $\log P(n) \sim a(n)^{\frac{1}{2}}$, we obtain by a simple calculation

$$\log f(x) \underset{x\to 1}{\sim} \frac{\pi^2}{6} \frac{\alpha}{1 - x}.$$

But

$$\log f(x) = \sum x^{a_i} + \tfrac{1}{2} \sum_i x^{2a_i} + \cdots = \sum_{k=1}^{\infty} b_k x^k.$$

Denote by $A(n)$ the number of a's not exceeding n. We have

$$B(n) = \sum_{k=1}^{n} b_k = \sum_{k=1}^{\infty} \frac{1}{k} A\left(\frac{n}{k}\right).$$

Thus

$$A(n) = \sum_{k=1}^{\infty} \frac{u(k)}{k} B\left(\frac{n}{k}\right).$$

But by the well known Tauberian theorem of Hardy-Littlewood,[11] we have

$$B(n) \sim \frac{\alpha\pi^2 n}{6}.$$

Hence

$$A(n) \sim \sum_{k=1}^{\infty} \frac{u(k)}{k^2} \frac{\alpha\pi^2 n}{6} \sim \alpha n. \quad \text{q.e.d.}$$

Similarly we can show that if $\log P'(n) = c[(\alpha/2)n]^{\frac{1}{2}}$, the density of the a's is α.

UNIVERSITY OF PENNSYLVANIA

[11] Hardy-Littlewood, ibid.

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

ON A PROBLEM OF I. SCHUR

BY P. ERDÖS AND G. SZEGÖ

(Received June 11, 1941)

TO THE MEMORY OF I. SCHUR

1. Introduction

1. Let $n \geqq 3$, and let Q_n denote the class of polynomials $f(x)$ of degree n satisfying the condition $|f(x)| \leqq 1$ in the interval $-1 \leqq x \leqq +1$. Let $Q_n(x_0)$ denote the subclass of Q_n characterized by the further restriction $f''(x_0) = 0$.

A well-known theorem of A. Markoff[1] states that $|f'(x)| \leqq n^2$ for $-1 \leqq x \leqq +1$ provided that $f(x) \in Q_n$; here $|f'(x)| = n^2$ holds if and only if $x = \pm 1$ and $f(x) = \pm T_n(x)$, where $T_n(x)$ denotes the n^{th} Tchebycheff polynomial. We observe that $T_n(x)$ does not belong to the classes $Q_n(\pm 1)$.

Some years ago I. Schur[2] proved the following interesting theorem: *Let* $-1 \leqq x_0 \leqq +1$, *and let* $f(x)$ *belong to* $Q_n(x_0)$. *Then* $|f'(x_0)| < \frac{1}{2}n^2$. Moreover he showed: *Let* m_n *be the least positive constant (depending only on* n*) such that* $|f'(x_0)| \leqq m_n \cdot n^2$ *for all* $f(x) \in Q_n(x_0)$, *and* x_0 *in* $-1 \leqq x \leqq +1$. *If* $\mu = \limsup_{n\to\infty} m_n$, *then*

$$0.217 \cdots \leqq \mu \leqq 0.465 \cdots . \tag{1.1}$$

Obviously

$$m_n \cdot n^2 = \max_{-1\leqq x_0\leqq +1} \max_{f(x)\,\in\, Q_n(x_0)} |f'(x_0)|. \tag{1.2}$$

The main purpose of the present note is to determine the constant μ and the polynomial $f(x)$ for which the extremum (1.2) is attained. In terms of the constant m_n, we obtain a bound for the derivative $f'(x)$ of a polynomial $f(x)$ which satisfies the condition that $|f'(x)|$ has a relative maximum at the point x considered.

2. *Let* $u_n(x)$ *be the polynomial of the class* $Q_n(+1)$ *for which* $u_n'(1)$ *is a maximum.* This polynomial $u_n(x) = u_n(x; A_n)$ can be determined from the transcendental equations (2.5), (2.6) and (2.17) of §2 (see below). It is a special case of a remarkable class of polynomials $u_n(x; A)$ considered first by G. Zolotareff[3]

---

[1] A. Markoff, *On a certain problem of D. I. Mendeleieff* (in Russian), Zapiski Imperatorskoi Akademii Nauk, vol. 62 (1889), pp. 1–24.

[2] I. Schur, *Über das Maximum des absoluten Betrages eines Polynoms in einem gegebenen Intervall*, Mathematische Zeitschrift, vol. 4 (1919), pp. 271–287.

[3] G. Zolotareff, (a) *On a question concerning a minimum value* (in Russian), Dissertation "pro venia legendi," published in lithographed form, 1868, Oeuvres, vol. 2 (1902), pp. 130–166; (b) *Application of elliptic functions to questions concerning functions which deviate the least from zero* (in Russian), Zapiski Imperatorskoi Akademii Nauk, vol. 30 (1877), Oeuvres, vol. 2, pp. 1–59; (c) *Sur l'application des fonctions elliptiques aux questions de maxima et minima*, Bulletin de l'Académie des Sciences de St.-Pétersbourg, series 3, vol. 24 (1878), pp. 305–310, Mélanges, 5, pp. 419–426, Oeuvres, vol. 1 (1931), pp. 369–374.

playing also a role in the important investigations of W. Markoff.[4] Recently N. Achyeser[5] used polynomials of the Zolotareff type in his investigations on polynomials of least deviation in two disjoint intervals. With the previous notation, our main result is:

THEOREM 1. *The extremum $m_n \cdot n^2$ in (1.2) is attained for $x_0 = +1$ and for the Zolotareff polynomials $\pm u_n(x)$ [or for $x_0 = -1$ and for $\pm u_n(-x)$], provided n is sufficiently large. Furthermore*

$$\lim_{n\to\infty} m_n = \mu \tag{1.3}$$

exists and

$$\mu = k^{-2}(1 - E/K)^2 = 0.3124\cdots, \tag{1.4}$$

where k^2 is the only root of the transcendental equation

$$(K - E)^3 + (1 - k^2)K - (1 + k^2)E = 0 \tag{1.5}$$

satisfying the condition $0 < k^2 < 1$. Here K and E are the complete elliptic integrals associated with the modulus k.

A further analysis and discussion of a few special cases furnishes the more informative

THEOREM 2. *If $n > 3$ the extremum $m_n \cdot n^2$ in (1.2) is attained in the cases mentioned in Theorem 1, and only in these cases. If $n = 3$, it is attained for $x_0 = 0$ and for the Tchebycheff polynomials $\pm T_3(x)$, and only then.*

In §§2 and 3 of the present paper we first study as a preparation the general polynomials $u_n(x; A)$ of Zolotareff and the special case $u_n(x) = u_n(x; A_n)$ mentioned above. The proof of Theorem 1 is then given in §§4 and 5, and that of Theorem 2 in §§6 and 7. In §8 we consider two problems of Zolotareff in which the polynomials $u_n(x; A)$ were first used; §9 contains another application.

The polynomials of Zolotareff occur in numerous other related extremum problems. They satisfy a simple differential equation by means of which they can be brought in relationship with the multiplication problem of elliptic integrals. In what follows we have tried to reduce the use of elliptic functions to a minimum.[6]

---

[4] W. Markoff, *Über Polynome, die in einem gegebenen Intervalle möglichst wenig von Null abweichen*, Mathematische Annalen, vol. 77 (1916), pp. 213–258. The Russian original appeared 1892.

[5] N. Achyeser, (a) *Über einige Funktionen, welche in zwei gegebenen Intervallen am wenigsten von Null abweichen*, Bulletin de l'Académie des Sciences de l'URSS, Classe des sciences mathématiques et naturelles, series 7, 1932, pp. 1163–1202; (b) *Über einige Funktionen, die in gegebenen Intervallen am wenigsten von Null abweichen*, Bulletin de la Société Physico-Mathématique de Kazan, series 3, vol. 3 (1928), pp. 1–69.

[6] Zolotareff and Achyeser make extensive use of the theory of elliptic functions; however, W. Markoff does not.

2. On the polynomials of Zolotareff

1. It is a classical fact that there is a unique polynomial $T_n(x)$ of degree n (the n^{th} polynomial of Tchebycheff) having the following property: The curve $y = T_n(x)$, $-1 \leqq x \leqq +1$, consists of n monotonic arcs varying between $+1$ and -1; $T_n(1) = 1$, and $T_n(-1) = (-1)^n$. This polynomial satisfies the differential equation

$$n^2(1 - y^2) = (1 - x^2)y'^2 \tag{2.1}$$

from which follows

$$y = \cos\left\{n \int_1^x (1 - t^2)^{-\frac{1}{2}}\, dt\right\}. \tag{2.2}$$

2. We show that there are infinitely many polynomials y of degree n possessing the following property: The curve y, $-1 \leqq x \leqq +1$, consists of $n - 1$ monotonic arcs varying between $+1$ and -1, $y = 1$ for $x = 1$, and $y = (-1)^{n-1}$ for $x = -1$. Such a curve necessarily has $n - 1$ roots in $-1 \leqq x \leqq +1$ and consequently one more outside this interval. If this additional root is >1, y satisfies a differential equation of the form

$$n^2(1 - y^2) = (1 - x^2)y'^2 \frac{(B - x)(C - x)}{(A - x)^2} \tag{2.3}$$

where $y' = 0$ for $x = A$, $y = 1$ for $x = B$, $y = -1$ for $x = C$, and $1 < A < B < C$. A similar differential equation holds if the additional root of y mentioned above is <-1. (The second case can be obtained from the first one by replacing x by $-x$.)

Solving the differential equation (2.3), we obtain

$$y = \cos\left\{n \int_1^x (A - t)(B - t)^{-\frac{1}{2}}(C - t)^{-\frac{1}{2}}(1 - t^2)^{-\frac{1}{2}}\, dt\right\}. \tag{2.4}$$

From the properties of y mentioned above we find

$$\int_{-1}^{+1} (A - t)(B - t)^{-\frac{1}{2}}(C - t)^{-\frac{1}{2}}(1 - t^2)^{-\frac{1}{2}}\, dt = (n - 1)\pi/n, \tag{2.5}$$

$$\int_{+1}^{B} (A - t)(B - t)^{-\frac{1}{2}}(C - t)^{-\frac{1}{2}}(t^2 - 1)^{-\frac{1}{2}}\, dt = 0, \tag{2.6}$$

$$\int_{B}^{C} (t - A)(t - B)^{-\frac{1}{2}}(C - t)^{-\frac{1}{2}}(t^2 - 1)^{-\frac{1}{2}}\, dt = \pi/n. \tag{2.7}$$

By a well-known application of Cauchy's theorem we see that the sum of the integrals (2.5) and (2.7) is π, so that (2.7) is a consequence of (2.5).

Conversely, if (2.5) and (2.6) hold, an easy discussion (encircling the singular points -1, $+1$, B, C) shows that (2.4) is an analytic function single-valued and regular in the whole finite x-plane. If $x \to \infty$ we find $y = O(|\, x\, |^n)$, so that y

must be a polynomial of degree n. Of course it satisfies the differential equation (2.3), and it has all the properties mentioned above.

For later purposes we note that

$$y' = n^2 \frac{(A-1)^2}{(B-1)(C-1)} \quad \text{at} \quad x = 1, \tag{2.8}$$

$$(-1)^n y' = n^2 \frac{(A+1)^2}{(B+1)(C+1)} \quad \text{at} \quad x = -1. \tag{2.9}$$

These values can be obtained from the differential equation (2.3).

3. Lemma 1. *Of the three quantities A, B, C $(1 < A < B < C)$ satisfying the two transcendental equations (2.5) and (2.6), any one can be prescribed arbitrarily provided that*

$$A > 1, \text{ or } B > 1, \text{ or } C > c_n = 1 + 2\alpha_n = 1 + 2\tan^2[\pi/(2n)] \tag{2.10}$$

respectively; the two others are then uniquely determined. As A increases monotonically from 1 to $+\infty$, B and C increase likewise from 1 to $+\infty$ and from c_n to $+\infty$, respectively.

Furthermore the values of $y, y', \cdots, y^{(n)}$ for a fixed x not less than one, and the values of $(-1)^n y, (-1)^{n-1} y', \cdots, y^{(n)}$ for a fixed x not greater than -1, are increasing functions of A.

The only exceptions are $y = 1$ for $x = 1$ and $(-1)^n y = -1$ for $x = -1$. In particular, the expressions (2.8) and (2.9) are respectively increasing and decreasing functions of A.

In order to prove this Lemma, let B denote a fixed value, greater than 1, and let C be variable, such that $C > B$; we define $A = A(C)$ by (2.6) so that $1 < A < B$. Then

$$\int_1^B \left(\frac{dA}{dC} - \frac{1}{2}\frac{A-t}{C-t}\right)(B-t)^{-\frac{1}{2}}(C-t)^{-\frac{1}{2}}(t^2-1)^{-\frac{1}{2}}\,dt = 0;$$

hence

$$\frac{dA}{dC} = \frac{1}{2}\frac{A-t_0}{C-t_0}, \quad \text{where} \quad 1 < t_0 < B.$$

Now consider the function $\lambda(C)$ defined by the left-hand member of (2.5), where $A = A(C)$. We find

$$\begin{aligned}\lambda'(C) &= \int_{-1}^{+1} \left(\frac{dA}{dC} - \frac{1}{2}\frac{A-t}{C-t}\right)(B-t)^{-\frac{1}{2}}(C-t)^{-\frac{1}{2}}(1-t^2)^{-\frac{1}{2}}\,dt \\ &= \int_{-1}^{+1} \left(\frac{1}{2}\frac{A-t_0}{C-t_0} - \frac{1}{2}\frac{A-t}{C-t}\right)(B-t)^{-\frac{1}{2}}(C-t)^{-\frac{1}{2}}(1-t^2)^{-\frac{1}{2}}\,dt < 0,\end{aligned}$$

so that $\lambda(C)$ is decreasing. Let $C \to B$; then from (2.6) we see that $A \to B$, so that

$$\lambda(C) \to \int_{-1}^{+1} (1 - t^2)^{-\frac{1}{2}}\, dt = \pi.$$

Since $\lim_{C\to\infty} \lambda(C) = 0$, the equation $\lambda(C) = (n-1)\pi/n$ has precisely one solution.[7]

4. Further let $p(x)$ and $q(x)$ be two special cases of (2.4) corresponding to the values A', B', C' and A'', B'', C'' of A, B, C, respectively. First suppose that $p'(1) < q'(1)$. Considering the polynomial $\delta(x) = p(x) - q(x)$ at the n points in $-1 \leqq x \leqq +1$ at which $p(x) = \pm 1$ and assuming that $\delta(x) \not\equiv 0$, a familiar argument furnishes the existence of $n-1$ distinct points $+1 > \eta_1 > \eta_2 > \cdots > \eta_{n-1} > -1$ such that $\delta'(\eta_1) > 0, \delta'(\eta_2) < 0, \cdots$. Furthermore $\delta'(1) < 0$, so that $\delta'(x)$ has $n-1$ roots (that is, all its roots) in $-1 < x < +1$. The same holds for $\delta''(x), \delta'''(x), \cdots$ so that $\delta(x) < 0, \delta'(x) < 0, \delta''(x) < 0, \cdots$ for $x \geqq 1$ [except that $\delta(1) = 0$], and also $(-1)^n\delta(x) < 0$, $(-1)^{n-1}\delta'(x) < 0$, $(-1)^{n-2}\delta''(x) < 0, \cdots$ for $x \leqq -1$ [except that $\delta(-1) = 0$]. From this we easily conclude that the relations $A' < A''$, $B' < B''$, $C' < C''$ hold for the constants corresponding to $p(x)$ and $q(x)$.

If $p'(1) = q'(1)$ the previous argument still holds good [unless $\delta(x) \equiv 0$], except that $\delta'(1) = 0$ so that the roots of $\delta'(x)$ are in $-1 < x \leqq +1$. Consequently $\delta'(x) < 0$ for $x > 1$. Interchanging $p(x)$ and $q(x)$ we obtain $\delta'(x) > 0$, $x > 1$, which is a contradiction; so that in this case $p(x) \equiv q(x)$, $A' = A''$, $B' = B''$, $C' = C''$.

From the previous considerations we conclude that B and C are increasing functions of A. It remains to calculate the limits of B and C as $A \to 1$ and $A \to +\infty$. In the former case, (2.6) shows that $B \to 1$, and from (2.5) we obtain $C \to c_n$ since the equation

$$\int_{-1}^{+1} (1 + t)^{-\frac{1}{2}}(\gamma - t)^{-\frac{1}{2}}\, dt = (n-1)\pi/n$$

has the unique solution $\gamma = c_n$. If $A \to +\infty$ it is obvious that $B \to +\infty$, $C \to +\infty$. This completes the proof of Lemma 1.

5. In what follows we denote the polynomial (2.4) [for which (2.5) and (2.6) hold] by $y = u_n(x; A)$. We note that, from (2.4) and (2.10),

$$\begin{aligned} u_n(x; +1) &= \lim_{A\to 1} u_n(x; A) \\ &= \cos\left\{n \int_1^x (1 + t)^{-\frac{1}{2}}(c_n - t)^{-\frac{1}{2}}\, dt\right\} = -T_n\left(\frac{x - \alpha_n}{1 + \alpha_n}\right). \end{aligned} \tag{2.11}$$

Hence $u_n'(+1; +1) = 0$. Also

$$u_n''(+1; +1) = -(1 + \alpha_n)^{-2} T_n''[\cos(\pi/n)] = -\tfrac{1}{4}n^2 \cot^2[\pi/(2n)]. \tag{2.12}$$

[7] These considerations require only slight modifications if we replace the right-hand members in (2.5) and (2.6) by $\nu\pi/n$ and $\pi - \nu\pi/n, 1 \leqq \nu \leqq n-1$. The resulting polynomials have been used for various purposes by Achyeser; see loc. cit.

Further, let $A \to +\infty$ so that $B \to +\infty$ and $C \to +\infty$. From (2.5)

$$(A - t_1)(B - t_1)^{-\frac{1}{2}}(C - t_1)^{-\frac{1}{2}} = (n-1)/n$$

where t_1 is a suitably chosen number between 0 and 1. Hence $A(BC)^{-\frac{1}{2}} \to 1 - 1/n$, so that, from (2.4),

$$u_n(x; +\infty) = \lim_{A\to\infty} u_n(x; A) = T_{n-1}(x). \tag{2.13}$$

Hence

$$u_n'(+1; +\infty) = (n-1)^2; \tag{2.14}$$

$$u_n''(+1; +\infty) = \tfrac{1}{3}n(n-1)^2(n-2). \tag{2.15}$$

Therefore, as A increases from 1 to $+\infty$, $u_n'(+1; A)$ increases from 0 to $(n-1)^2$, and $u_n''(+1; A)$ increases from the negative value (2.12) to the positive value (2.15), corresponding respectively to $A = 1$ and $A = +\infty$. There is precisely one value of A, $A = A_n$, for which $u_n''(+1; A_n) = 0$. We denote the corresponding values of B and C by B_n and C_n. In §§4 and 5 we shall prove that the function $u_n(x; A_n)$ furnishes the solution of I. Schur's problem formulated above, provided n is sufficiently large.

From the differential equation (2.3) we obtain

$$u_n''(+1; A) = \tfrac{1}{3}n^2 \frac{(A-1)^2}{(B-1)(C-1)}\left\{n^2 \frac{(A-1)^2}{(B-1)(C-1)} - 1 - 2\left(\frac{2}{A-1} - \frac{1}{B-1} - \frac{1}{C-1}\right)\right\}, \tag{2.16}$$

so that the condition $u_n''(+1; A) = 0$ is equivalent to

$$n^2 \frac{(A-1)^2}{(B-1)(C-1)} = 1 + 2\left(\frac{2}{A-1} - \frac{1}{B-1} - \frac{1}{C-1}\right). \tag{2.17}$$

The transcendental equations (2.5), (2.6) and (2.17) determine the constants $A = A_n$, $B = B_n$, $C = C_n$ uniquely. These constants depend only on n.

The polynomial $y = u_n(x; A_n)$ is completely determined by the following conditions: The curve y, $-1 \leqq x \leqq +1$, consists of $n-1$ monotonic arcs varying between $+1$ and -1, $y = 1$ for $x = 1$, $y = (-1)^{n-1}$ for $x = -1$ and $y'' = 0$ for $x = 1$.

3. The limiting process $n \to \infty$

1. First we prove the following

LEMMA 2. *The constants A_n, B_n, C_n defined by the transcendental equations* (2.5), (2.6), (2.17) *satisfy*

$$\lim_{n\to\infty} n^2(A_n - 1) = a^2/2, \qquad \lim_{n\to\infty} n^2(B_n - 1) = b^2/2, \qquad \lim_{n\to\infty} n^2(C_n - 1) = c^2/2 \tag{3.1}$$

where $0 < a < b < c$. The numerical values of a, b, c are given in (3.17).

By Lemma 1

$$\liminf_{n\to\infty} n^2(C_n - 1) \geqq \pi^2/2$$

and from (2.17)

$$n^2\frac{A_n - 1}{C_n - 1} \geqq \frac{2}{A_n - 1} - \frac{2}{C_n - 1},$$

$$n^4(A_n - 1)^2 + 2n^2(A_n - 1) \geqq 2n^2(C_n - 1), \tag{3.2}$$

so that

$$\liminf_{n\to\infty} n^2(A_n - 1) \geqq (1 + \pi^2)^{\frac{1}{2}} - 1. \tag{3.3}$$

The same inequality holds if we replace A_n by B_n.

On the other hand, let us assume that $n^2(C_n - 1) \to +\infty$ for a proper subsequence $n = n_\nu$ as $\nu \to \infty$; then, from (3.2), $n^2(A_n - 1) \to +\infty$, so that $n^2(B_n - 1) \to +\infty$. Therefore, by (2.17), for the same subsequence $n = n_\nu$,

$$\frac{(A_n - 1)^2}{(B_n - 1)(C_n - 1)} \to 0, \tag{3.4}$$

Now let ω be a fixed positive number; for large n, from (2.5),

$$\pi = n\int_{-1}^{+1} (1 - t^2)^{-\frac{1}{2}}\, dt\{1 - (A_n - t)(B_n - t)^{-\frac{1}{2}}(C_n - t)^{-\frac{1}{2}}\}$$

$$> n\int_0^{\omega/n} d\varphi\{1 - (A_n - \cos\varphi)(B_n - \cos\varphi)^{-\frac{1}{2}}(C_n - \cos\varphi)^{-\frac{1}{2}}\} \tag{3.5}$$

$$= \int_0^{\omega} d\psi\{1 - (A_n - \cos(\psi/n))(B_n - \cos(\psi/n))^{-\frac{1}{2}}(C_n - \cos(\psi/n))^{-\frac{1}{2}}\}.$$

Since

$$(A_n - \cos(\psi/n))(B_n - \cos(\psi/n))^{-\frac{1}{2}}(C_n - \cos(\psi/n))^{-\frac{1}{2}}$$

$$\leqq (A_n - 1)(B_n - 1)^{-\frac{1}{2}}(C_n - 1)^{-\frac{1}{2}}\left\{1 + \frac{1 - \cos(\psi/n)}{A_n - 1}\right\}$$

and since for $n = n_\nu$, as $\nu \to \infty$,

$$\frac{n^2(1 - \cos(\psi/n))}{n^2(A_n - 1)} \to 0$$

uniformly in ψ, for $0 \leqq \psi \leqq \omega$, we find $\pi \geqq \omega$. This is a contradiction if we choose $\omega > \pi$. Thus we have proved that the points of accumulation of the sequences $n^2(A_n - 1)$, $n^2(B_n - 1)$, $n^2(C_n - 1)$ are positive and finite.

2. Now let $n = n_\nu$ be a subsequence for which the limits (3.1) exist, where

$0 < a \leqq b \leqq c < +\infty$. From (2.5), (2.6) and (2.7) we shall derive $a < b < c$ and

$$\int_0^\infty \{1 - (a^2 + u^2)(b^2 + u^2)^{-\frac{1}{2}}(c^2 + u^2)^{-\frac{1}{2}}\}\, du = \pi, \tag{3.6}$$

$$\int_0^b (a^2 - u^2)(b^2 - u^2)^{-\frac{1}{2}}(c^2 - u^2)^{-\frac{1}{2}}\, du = 0, \tag{3.7}$$

$$\int_b^c (u^2 - a^2)(u^2 - b^2)^{-\frac{1}{2}}(c^2 - u^2)^{-\frac{1}{2}}\, du = \pi. \tag{3.8}$$

Also from (2.17) by the same limiting process ($n = n_\nu$, $\nu \to \infty$),

$$\frac{a^4}{b^2 c^2} - 4\left(\frac{2}{a^2} - \frac{1}{b^2} - \frac{1}{c^2}\right) = 0. \tag{3.9}$$

Instead of (3.7) we can show more precisely

$$\begin{cases} n\displaystyle\int_1^{A_n} (A_n - t)(B_n - t)^{-\frac{1}{2}}(C_n - t)^{-\frac{1}{2}}(t^2 - 1)^{-\frac{1}{2}}\, dt \\ \qquad\qquad \to \displaystyle\int_0^a (a^2 - u^2)(b^2 - u^2)^{-\frac{1}{2}}(c^2 - u^2)^{-\frac{1}{2}}\, du, \\ n\displaystyle\int_{A_n}^{B_n} (t - A_n)(B_n - t)^{-\frac{1}{2}}(C_n - t)^{-\frac{1}{2}}(t^2 - 1)^{-\frac{1}{2}}\, dt \\ \qquad\qquad \to \displaystyle\int_a^b (u^2 - a^2)(b^2 - u^2)^{-\frac{1}{2}}(c^2 - u^2)^{-\frac{1}{2}}\, du, \end{cases} \tag{3.10}$$

the two limits being the same.

First, (3.9) is obvious and this equation shows that $a = b = c$ is impossible. In case $a < b = c$ both formulas (3.10) follow easily [writing $t = 1 + u^2/(2n^2)$]; but the first limit is finite and the second one turns out to be $+\infty$, which is a contradiction. In case $a = b < c$ the same formulas can be easily established again, but the first limit is positive whereas the second one is 0 [since $\max\{(t - A_n)(C_n - t)^{-\frac{1}{2}}(t^2 - 1)^{-\frac{1}{2}}\}$, $A_n \leqq t \leqq B_n$, is bounded]. Therefore $a < b < c$.

Now (3.7) and (3.8) follow directly, and (3.6) can also be easily obtained. However (3.6) follows also from (3.8) by applying Cauchy's theorem to

$$f(z) = 1 - (a^2 - z^2)(b^2 - z^2)^{-\frac{1}{2}}(c^2 - z^2)^{-\frac{1}{2}}$$

integrated along the half-circle $|z| = R$, $\Re z \geqq 0$ and along the segment $\Re z = 0$, $-R \leqq \Im z \leqq +R$, $R \to +\infty$.

3. Substituting $u^2 = b^2 \sin^2\varphi$ in (3.7) and $u^2 = c^2 - (c^2 - b^2)\sin^2\varphi$ in (3.8) we find

$$\int_0^{\pi/2} (a^2 - b^2\sin^2\varphi)(c^2 - b^2\sin^2\varphi)^{-\frac{1}{2}}\, d\varphi = 0, \tag{3.11}$$

$$\int_0^{\pi/2} \{c^2 - a^2 - (c^2 - b^2)\sin^2\varphi\}\{c^2 - (c^2 - b^2)\sin^2\varphi\}^{-\frac{1}{2}}\,d\varphi = \pi. \tag{3.12}$$

Using the standard notation these equations can be written in the form

$$(1 - a^2/c^2)K = E, \qquad cE' - (a^2/c)K' = \pi \tag{3.13}$$

where the complete elliptic integrals K and E belong to the modulus $k = b/c$. Eliminating a^2/c^2 we find

$$E/K + (E' - \pi/c)/K' = 1. \tag{3.14}$$

Comparing this with the classical equation[8]

$$EK' + E'K - KK' = \pi/2 \tag{3.15}$$

we obtain $c = 2K$. Hence

$$a^2 = 4K(K - E), \qquad b = 2kK, \qquad c = 2K. \tag{3.16}$$

The relation (3.9) furnishes the transcendental equation (1.5) of Theorem 1 (see §1) for the modulus k. This equation has precisely one root as k^2 goes from 0 to 1 [which shows that the limits (3.1) exist as $n \to \infty$ unrestrictedly]. Indeed, differentiating the left-hand member of (1.5) with respect to k^2,[9] we have

$$\tfrac{3}{2}E\{k'^{-2}(K - E)^2 - 1\},$$

where k' is the complementary modulus. The expression in the curly bracket increases with k^2, as the well-known power series expansion of K and E shows; it is negative for small k^2 and positive as k^2 approaches 1. Therefore the left-hand member of (1.5) first decreases and then increases; but for $k^2 = 0$ it is zero and for $k^2 \to 1 - 0$ it tends to $+\infty$. This establishes Lemma 2.

Using the tables of Milne-Thomson[10] we find

$$k^2 = 0.84\cdots, \quad a^2 = 11.4055\cdots, \quad b = 4.3245\cdots, \quad c = 4.7185\cdots, \quad a^4/b^2c^2 = 0.3124\cdots. \tag{3.17}$$

We also note that (2.4) implies that

$$\lim_{n\to\infty} u_n(\cos(z/n); A_n) = \cos\left\{\int_0^z (a^2 + u^2)(b^2 + u^2)^{-\frac{1}{2}}(c^2 + u^2)^{-\frac{1}{2}}\,du\right\} \tag{3.18}$$

uniformly in z, for all complex z such that $|z| \leqq R$.

4. Another limiting formula important for the proof of Theorem 1, is

LEMMA 3. *Let $A = A_n'$ be a sequence of values such that $A_n' - 1 = o(n^{-2})$.*

[8] See for instance, E. T. Whittaker and G. N. Watson, *A course of Modern Analysis*, Fourth edition, 1935, p. 520.

[9] See Whittaker-Watson, loc. cit. p. 521.

[10] L. M. Milne-Thomson, *Ten-figure table of the complete elliptic integrals K, K', E, E' and a table of $1/\vartheta_3^2(0|\tau)$, $1/\vartheta_3^2(0|\tau')$*, Proceedings of the London Mathematical Society, series 2, vol. 33 (1932), pp. 160–164.

Denoting the corresponding values of B and C determined from the equations (2.5) *and* (2.6) *by B_n' and C_n', respectively, we have*

$$\lim_{n\to\infty} u_n(\cos(z/n); A_n') = -\cos\{(\pi^2 + z^2)^{\frac{1}{2}}\}. \tag{3.19}$$

The last equation holds uniformly in z, for all complex z such that $|z| \leqq R$.

We note that (3.19) arises from (3.18) on writing $a = b = 0$, $c = \pi$.

For the proof we use an argument similar to that of Part 1. Let ω be fixed, $\omega > 0$; we find [see (3.5)]

$$\pi > \int_0^{\omega} d\psi\{1 - (A_n' - \cos(\psi/n))(B_n' - \cos(\psi/n))^{-\frac{1}{2}}(C_n' - \cos(\psi/n))^{-\frac{1}{2}}\}. \tag{3.20}$$

Assuming for a certain subsequence $n = n_\nu$, $\nu \to \infty$, that the limits

$$\lim n^2(B_n' - 1) = \beta, \qquad \lim n^2(C_n' - 1) = \gamma$$

exist, we have $\beta \geqq 0$, $\gamma \geqq \pi^2/2$. Thus we conclude from (3.20)

$$\pi \geqq \int_0^{\omega} d\psi\{1 - (\psi^2/2)(\beta + \psi^2/2)^{-\frac{1}{2}}(\gamma + \psi^2/2)^{-\frac{1}{2}}\},$$

so that

$$\pi \geqq \int_0^{\infty} d\psi\{1 - (\psi^2/2)(\beta + \psi^2/2)^{-\frac{1}{2}}(\gamma + \psi^2/2)^{-\frac{1}{2}}\}. \tag{3.21}$$

Now

$$\pi = \int_0^{\infty} d\psi\{1 - \psi(\pi^2 + \psi^2)^{-\frac{1}{2}}\}; \tag{3.22}$$

consequently (3.21) and (3.22) involve a contradiction, unless $\beta = 0$, $\gamma = \pi^2/2$.

Further

$$u_n(\cos(z/n); A_n') = \cos\left\{\int_0^z (A_n' - \cos(\psi/n))(B_n' - \cos(\psi/n))^{-\frac{1}{2}}(C_n' - \cos(\psi/n))^{-\frac{1}{2}}\, d\psi\right\}. \tag{3.23}$$

Now let $0 < \epsilon < \pi < R$ and $|z| = R$. Then

$$\int_0^{\epsilon} (A_n' - \cos(\psi/n))(B_n' - \cos(\psi/n))^{-\frac{1}{2}}(C_n' - \cos(\psi/n))^{-\frac{1}{2}}\, d\psi$$

$$\leqq \int_0^{\epsilon} (A_n' - \cos(\psi/n))^{\frac{1}{2}}(C_n' - \cos(\psi/n))^{-\frac{1}{2}}\, d\psi \to \int_0^{\epsilon} \psi(\pi^2 + \psi^2)^{-\frac{1}{2}}\, d\psi$$

as $n \to \infty$; the last integral is arbitrarily small with ϵ. Integrating from ϵ to z, we can assume that $\psi \neq 0, \pm\pi$ on the path of integration; and the assertion follows immediately from (3.23) for $n \to \infty$.

4. Proof of Theorem 1

In what follows, the symbols Q_n, $Q_n(x_0)$ defined in §1 are used.

1. LEMMA 4. *Suppose* $-1 \leqq x_0 \leqq +1$, *and let* $f_0(x)$ *be a polynomial of the class* $Q_n(x_0)$ *for which* $\max |f'(x_0)|$, $f(x) \in Q_n(x_0)$, *is attained. Then* $|f_0(x)|$ *assumes its maximum* 1 *at least* n *times in* $-1 \leqq x \leqq +1$.

The proof follows the usual lines. Let $f_0'(x_0) > 0$ and let us suppose that the assertion of Lemma 4 does not hold. Denote by $x_1, x_2, \cdots, x_l$; $l < n$, the distinct values in $-1 \leqq x \leqq +1$ for which $|f_0(x_\nu)| = 1$ and write $\omega(x) = \prod_{\nu=1}^{l}(x - x_\nu)$. If $-1 < x_0 < +1$ we have $x_0 \neq x_\nu$ [otherwise $f_0'(x_0)$ would be 0]. However if $x_0 = \pm 1$ we may have $x_0 = x_\nu$, in which case $\omega(x_0) = 0$ but $\omega'(x_0) \neq 0$.

We form the polynomial

$$(4.1) \qquad r(x) = -\sum_{\nu=1}^{l} \operatorname{sgn} f_0(x_\nu) \frac{\omega(x)}{\omega'(x_\nu)(x - x_\nu)} + \omega(x)\{a(x - x_0) + b\}$$

and want to determine the constants a and b such that $r'(x_0) > 0$, $r''(x_0) = 0$; this can certainly be done provided the linear equations

$$a\omega(x_0) + b\omega'(x_0) = G,$$

$$2a\omega'(x_0) + b\omega''(x_0) = H$$

have a determinant $\neq 0$. Now $\omega(x_0)\omega''(x_0) - 2\{\omega'(x_0)\}^2 \neq 0$ is obvious if $\omega(x_0) = 0$ (cf. above); but if $\omega(x_0) \neq 0$,

$$\frac{\omega''(x_0)}{\omega(x_0)} - 2\left\{\frac{\omega'(x_0)}{\omega(x_0)}\right\}^2 = -\sum_{\nu=1}^{l}(x_0 - x_\nu)^{-2} - \left\{\frac{\omega'(x_0)}{\omega(x_0)}\right\}^2 < 0.$$

Obviously $r(x)$ is of degree $l + 1 \leqq n$ and we find for sufficiently small $\epsilon > 0$ that $|f_0(x) + \epsilon r(x)| \leqq 1$ in $-1 \leqq x \leqq +1$; hence $f_0(x) + \epsilon r(x)$ belongs to $Q_n(x_0)$. On the other hand $f_0'(x_0) + \epsilon r'(x_0) > f_0'(x_0)$ which is a contradiction. This proves Lemma 4.

2. Let the extremum (1.2) be attained for the value x_0 and for $f(x) = f_0(x)$, $f_0(x) \in Q_n(x_0)$. Then $f_0''(x_0) = 0$, and $f_0(x)$ possesses the property formulated in Lemma 4. Further we show that $f_0'''(x_0) \neq 0$. By Lemma 4, $|f_0(x)|$ attains its relative maximum 1 in $-1 < x < +1$ for at least $n - 2$ distinct points for which $f_0'(x) = 0$. Since $f_0''(x)$ vanishes an odd number of times between two consecutive roots of $f_0'(x)$, we find that $f_0''(x)$ has precisely one simple root between two consecutive roots of $f_0'(x)$, and these roots of $f_0''(x)$ are maximum points of $|f_0'(x)|$. The number of these maximum points is at least $n - 3$. If x_0 is one of these points, we must have $f_0'''(x_0) \neq 0$. If x_0 is different from these maximum points (whose number in this case is $n - 3$), then we must have again $f_0'''(x_0) \neq 0$, and thus there is a relative maximum of $|f_0'(x)|$ at $x = x_0$.

If we assume that $f_0'(x_0) > 0$ then $f_0''(x_0) = 0$, $f_0'''(x_0) < 0$, so that $f_0'(x)$ has a relative maximum at $x = x_0$.

Now we distinguish various cases.

(a) $x_0 = \pm 1$.

Let $x_0 = +1$ and let us denote an extremum polynomial of our problem by $u_n(x)$, $u_n'(1) > 0$, $u_n''(1) = 0$. As we showed before, $u_n'(x)$ has at least $n-2$ and $u_n''(x)$ at least $n-3$ distinct roots in $-1 < x < +1$. Since $u_n''(1) = 0$, we find that $n-2$ is the precise number of roots of $u_n'(x)$ in $-1 < x < +1$. Consequently $|u_n(-1)| = |u_n(+1)| = 1$; and, since $u_n'(1) > 0$, we find $u_n(1) = 1$, $u_n(-1) = (-1)^{n-1}$.

Thus the curve $y = u_n(x)$, $-1 \leqq x \leqq +1$, consists of $n-1$ monotonic arcs varying between $+1$ and -1, and $u_n(1) = 1$, $u_n(-1) = (-1)^{n-1}$, $u_n'(1) > 0$, $u_n''(1) = 0$.

Hence from the last remark of §2 we conclude that $u_n(x)$ is identical with the polynomial $u_n(x; A_n)$ defined there.

Consequently, under the assumption $x_0 = \pm 1$, the extremum polynomials of our problem are $\pm u_n(x; A_n)$ and $\pm u_n(-x; A_n)$, respectively. The asymptotic value of $|u_n'(1; A_n)|$ is $a^4b^{-2}c^{-2}\cdot n^2$ [see (3.17)].

(b) $-1 < x_0 < +1$, and there exists a polynomial $g(x)$ of Q_n for which $|g'(x_0)| > |f_0'(x_0)|$. Suppose $f_0'(x_0) > 0$, $g'(x_0) > 0$.

Consider the polynomial $h_\epsilon(x) = f_0(x) + \epsilon\{g(x) - f_0(x)\}$, $0 < \epsilon < 1$. Obviously $h_\epsilon(x) \in Q_n$; furthermore $h_\epsilon'(x_0) > f_0'(x_0)$. For sufficiently small ϵ there is a root of $h_\epsilon''(x)$ in the neighborhood of x_0, x_0' say, and $h_\epsilon'(x)$ attains a positive relative maximum at $x = x_0'$. We evidently have

$$h_\epsilon'(x_0') \geqq h_\epsilon'(x_0) > f_0'(x_0)$$

which shows that $f_0(x)$ cannot be the extremum polynomial.

5. Proof of Theorem 1 (continued)

The remaining case requires a more elaborate discussion. This case is:

(c) $-1 < x_0 < +1$ and $f_0(x)$ is the polynomial in Q_n with the maximum value of $f'(x_0)$.

Then W. Markoff has shown[11] that $f_0(x)$ must be one of the polynomials

$$\pm T_n(x), \qquad \pm T_{n-1}(x), \qquad \pm T_n\left(\frac{x-\alpha}{1+\alpha}\right), \qquad \pm T_n\left(\frac{x+\alpha}{1+\alpha}\right), \tag{5.1}$$

$$\pm u_n(\pm x; A)$$

where $0 < \alpha < \alpha_n = \tan^2[\pi/(2n)]$, and $u_n(x; A)$ are the Zolotareff polynomials defined and discussed above. As $n \to \infty$, the largest relative maximum of $|T_n(x)|$ in $-1 < x < +1$ is asymptotically $M\cdot n^2$ where $-M$ is the minimum of $\sin\theta/\theta$ for real θ, that is $M = 0.2172\cdots$. Comparing this result with the asymptotic value of $u_n'(1; A_n)$, that is with $a^4b^{-2}c^{-2}\cdot n^2$ [see (3.17)], we see that for large values of n the four first types in (5.1) can be excluded.

[11] Loc. cit. p. 249.

As W. Markoff has further shown[12], $f_0(x) = \pm u_n(x; A)$ if and only if (a) x_0 belongs to certain open intervals in $-1 < x < +1$, and (b):

$$\frac{d}{dx}\left\{\frac{(1-x^2)u_n'(x;A)}{x-A}\right\} = 0 \qquad \text{at} \quad x = x_0. \tag{5.2}$$

Since $u_n'(x_0\,;A) \neq 0$, and $u_n''(x_0\,;A) = 0$, the latter-mentioned condition implies that

$$x_0 = A - (A^2-1)^{\frac{1}{2}}, \qquad A - 1 = (1-x_0)^2/(2x_0), \tag{5.3}$$

so that $0 < x_0 < +1$. Now we distinguish again two cases:

(c′) $0 < x_0 \leqq (1 - 16n^{-2})^{\frac{1}{2}}$. According to S. Bernstein's theorem

$$|\,u_n'(x\,;A)\,| \leqq n(1-x^2)^{-\frac{1}{2}} \leqq n^2/4. \tag{5.4}$$

(c″) $(1 - 16n^{-2})^{-\frac{1}{2}} < x_0 < 1$. Then $A - 1 = A_n' - 1 = O(n^{-4})$. Now we assume that this case occurs for an infinite number of values of n, and we write $x_0 = \cos(z_0/n)$; then z_0 is bounded. From Lemma 3 we conclude that

$$\lim n^{-2}u_n'\,(\cos(z/n);\,A_n') = -\frac{\sin\{(\pi^2+z^2)^{\frac{1}{2}}\}}{(\pi^2+z^2)^{\frac{1}{2}}}. \tag{5.5}$$

The maximum of the absolute value of the last expression for real z is $M = 0.2172\cdots$ so that this case can be also eliminated.

The assumption $f_0(x) = \pm u_n(-x; A)$ can be dealt with similarly.

Thus for large n only Case (a) remains. This completes the proof of Theorem 1.

6. Proof of Theorem 2

1. First we consider again the case (c) defined in §5 and let x_0 belong to one of the open intervals in $-1 \leqq x \leqq +1$ in which the maximum of $f'(x_0), f(x) \in Q_n$, is attained for the Zolotareff polynomial $f(x) = u_n(x; A)$. [The argument is similar for $-u_n(x; A)$ or $\pm u_n(-x; A)$.] Then $f(x) = u_n(x; A) = f_0(x)$, where $f_0(x)$ has the same meaning as in §§4 and 5, so that $f_0(x) \in Q_n(x_0)$; that is, $f_0''(x_0) = 0$. We have $f_0'(x_0) > 0$, $f_0'''(x_0) < 0$.

By an important theorem of W. Markoff[13], to every positive ϵ correspond values x_1 such that[14]

(α) $0 < |\,x_1 - x_0\,| < \epsilon$;

(β) if $f_1(x) = u_n(x; A')$ denotes the polynomial of Q_n for which $f'(x_1)$ becomes a maximum, then

$$f_1'(x_1) > f_0'(x_0). \tag{6.1}$$

---

[12] Loc. cit. pp. 233–246.

[13] Loc. cit. p. 257.

[14] In fact, a whole half-neighborhood of x_0 satisfies this condition.

Now if ϵ is sufficiently small, $f_1''(x)$ will have a root, say x_1', in the neighborhood of x_0; we can assume that $-1 < x_1' < +1$. Also $f_1'''(x_1') < 0$, so that $f_1'(x)$ has a relative maximum at $x = x_1'$; hence

$$f_1'(x_1') \geqq f_1'(x_1) > f_0'(x_0), \tag{6.2}$$

which shows that $f_0(x)$ can not be the solution of our problem.

This argument leaves as the only possibilities for $f_0(x)$ either the Zolotareff polynomials $\pm u_n(x; A_n)$ with $x_0 = \pm 1$, or the Tchebycheff polynomials $\pm T_n(x)$.

2. Let D_n be the largest root of $u_n(x; A_n)$, $B_n < D_n < C_n$. Using the convexity of $u_n(x; A_n)$ for $x > 1$, we deduce

$$D_n - B_n > C_n - D_n. \tag{6.3}$$

Further we make use of a theorem of I. Schur on the largest roots of the derivatives of an algebraic equation with only real roots.[15] Applying this theorem to $u_n(x; A_n)$ we obtain

$$D_n - A_n \leqq A_n - 1 \tag{6.4}$$

so that

$$2(A_n - 1) \geqq D_n - 1 > \tfrac{1}{2}(B_n - 1 + C_n - 1) > \{(B_n - 1)(C_n - 1)\}^{\frac{1}{2}}.$$

Hence, from (2.8),

$$u_n'(1; A_n) > n^2/4. \tag{6.5}$$

3. On the other hand we show that

$$|T_n'(x)| \leqq n^2/4 \quad \text{if} \quad T_n''(x) = 0 \tag{6.6}$$

provided $n \geqq 5$ (with equality only if $n = 5$). Incidentally, I. Schur has proved (6.6) for all large n.[16]

Let φ be a root of the equation $\tan n\varphi = n \tan \varphi$, $0 < \varphi < \pi/2$. Then the assertion is

$$n\left|\frac{\sin n\varphi}{\sin \varphi}\right| = n^2(n^2 \sin^2 \varphi + \cos^2 \varphi)^{-\frac{1}{2}} \leqq n^2/4, \qquad \sin \varphi \geqq \left(\frac{15}{n^2 - 1}\right)^{\frac{1}{2}}. \tag{6.7}$$

It is sufficient to show this for the largest root $x_n = \cos \varphi_n$ of $T_n''(x)$; that is, for the smallest positive value φ_n, $\pi < n\varphi_n < 3\pi/2$, satisfying the equation above.

The function

$$h(\psi) = \frac{\tan n\psi}{n \tan \psi} \tag{6.8}$$

---

[15] I. Schur, *Zwei Sätze über algebraische Gleichungen mit lauter reellen Wurzeln*, Journal für die reine und angewandte Mathematik, vol. 144 (1914), pp. 75–88.

[16] I. Schur, loc. cit.[2], p. 277.

increases from 0 to $+\infty$ as ψ increases from π/n to $3\pi/(2n)$. Let z be the smallest positive root of the equation $\tan z = z$, $\pi < z < 3\pi/2$. Since

$$h(z/n) = \frac{z}{n \tan (z/n)} < 1, \tag{6.9}$$

we have $\varphi_n > z/n$, so that (6.7) follows from

$$\sin (z/n) \geqq \left(\frac{15}{n^2 - 1}\right)^{\frac{1}{2}}. \tag{6.10}$$

Since $n \sin (z/n)$ increases and $n^2/(n^2 - 1)$ decreases as n increases, the last inequality will be proved for $n \geqq 6$ if we prove it for $n = 6$. But

$$\sin (z/6) \geqq (3/7)^{\frac{1}{2}} = 0.6546 \cdots, \tag{6.11}$$

since[17] $z = 4.4934 \cdots$ and $\sin (z/6) = 0.6808 \cdots$.

In the case $n = 5$ we have

$$T_5''(x) = 320x^3 - 120x, \quad x_5 = \cos \varphi_5 = (3/8)^{\frac{1}{2}}, \quad \sin \varphi_5 = (5/8)^{\frac{1}{2}}. \tag{6.12}$$

Comparing (6.5) and (6.6) we obtain $\pm u_n(\pm x; A_n)$ as the only eligible extremum polynomials [and $x_0 = \pm 1$ as the points at which the extremum is obtained] provided $n \geqq 5$.

7. Proof of Theorem 2 (continued)

The previous result holds also for $n = 4$, as a direct discussion shows; however, it fails for $n = 3$.

1. We have for $n = 4$:

$$T_4(x) = 8x^4 - 8x^2 + 1, \quad T_4'(x) = 32x^3 - 16x, \quad T_4''(x) = 96x^2 - 16, \tag{7.1}$$

so that, with the same notation as before, $x_4 = 6^{-\frac{1}{2}}$ and

$$| T_4'(x_4) | = (16/3)(2/3)^{\frac{1}{2}} = 4.3546 \cdots. \tag{7.2}$$

On the other hand, let us denote by y_1 and y_2, the values of x for which the relative extrema of $u_4(x; A_4)$ in $-1 \leqq x \leqq +1$ are attained; thus $-1 < y_1 < y_2 < +1$, say. Then

$$u_4(x; A_4) = 1 - \lambda(1 - x)(B_4 - x)(y_1 - x)^2 \tag{7.3}$$

must satisfy the following conditions:

$$\begin{cases} (\alpha): & u_4(-1; A_4) = -1, & \lambda(B_4 + 1)(y_1 + 1)^2 = 1, \\ (\beta): & u_4(y_2; A_4) = -1, & \lambda(1 - y_2)(B_4 - y_2)(y_1 - y_2)^2 = 2, \\ (\gamma): & u_4'(y_2; A_4) = 0, & \dfrac{1}{y_2 - 1} + \dfrac{1}{y_2 - B_4} + \dfrac{2}{y_2 - y_1} = 0, \\ (\delta): & u_4''(1; A_4) = 0, & 2B_4 + y_1 = 3. \end{cases} \tag{7.4}$$

[17] See, for instance, E. Jahnke-F. Emde, Funktionentafeln, 1933, p. 30.

Hence $B_4 < 2$. Let

$$1 - y_2 = h(B_4 - 1), \qquad B_4 - y_2 = (h+1)(B_4 - 1), \qquad y_2 - y_1 = (2-h)(B_4 - 1); \tag{7.5}$$

then (γ) becomes:

$$\frac{1}{h} + \frac{1}{h+1} + \frac{2}{h-2} = 0; \quad \text{i.e.,} \quad h = (1 + (33)^{\frac{1}{2}})/8 = 0.8430\cdots. \tag{7.6}$$

Further, writing $v(x) = x(x+1)(x-2)^2$, we obtain from (α) and (β)

$$v\left(\frac{2}{B_4 - 1}\right) = v(h). \tag{7.7}$$

Since $v(x) = v(h)$ has h as a double root, it can be reduced to a quadratic equation giving

$$\frac{2}{B_4 - 1} = 3/2 - h + \tfrac{1}{2}(10h + 5)^{\frac{1}{2}} = 2.4893\cdots. \tag{7.8}$$

Now

$$u_4'(1; A_4) = \lambda(B_4 - 1)(1 - y_1)^2 = 4\,\frac{\dfrac{2}{B_4 - 1}}{v\left(\dfrac{2}{B_4 - 1}\right)} = 4.7881\cdots. \tag{7.9}$$

Comparison of this value with (7.2) furnishes $u_4(x; A_4)$ as the solution.

2. Finally in the case $n = 3$,

$$T_3(x) = 4x^3 - 3x, \qquad T_3'(x) = 12x^2 - 3, \qquad T_3''(x) = 24x, \qquad x_3 = 0, \quad |T_3'(x_3)| = 3. \tag{7.10}$$

On the other hand,

$$u_3(x; A_3) = 1 - \lambda(1 - x^2)(B_3 - x) \tag{7.11}$$

with a relative minimum at $x = y_1$, $-1 < y_1 < +1$, satisfies the following conditions:

$$\begin{aligned} &(\alpha)\colon\ u_3(y_1; A_3) = -1, && \lambda(1 - y_1^2)(B_3 - y_1) = 2, \\ &(\beta)\colon\ u_3'(y_1; A_3) = 0, && 3y_1^2 - 2B_3 y_1 - 1 = 0, \\ &(\gamma)\colon\ u_3''(1; A_3) = 0, && B_3 = 3, \end{aligned} \tag{7.12}$$

so that

$$\begin{cases} y_1 = 1 - 2\cdot 3^{-\frac{1}{2}}, \qquad \lambda = 3^{\frac{1}{2}}/8, \\ u_3(x; A_3) = 1 - 3^{\frac{1}{2}}(1 - x^2)(3 - x)/8, \quad u_3'(1; A_3) = 3^{\frac{1}{2}}/2 < 3. \end{cases} \tag{7.13}$$

This completes the proof of Theorem 2.

8. Two problems of Zolotareff

1. The previous considerations permit a very simple approach to the following interesting theorem of Zolotareff:[18]

THEOREM 3. *Let σ be a given positive number and $f(x)$ an arbitrary polynomial of degree n of the form*

$$f(x) = x^n - \sigma x^{n-1} + \cdots . \tag{8.1}$$

Then $\max |f(x)|$, $-1 \leqq x \leqq +1$, *is minimized if and only if*

(a) $f(x) = \text{const.}\, u_n(x; A)$ *provided* $\sigma \geqq n\alpha_n$,

(b) $f(x) = 2^{1-n}(1 + \sigma/n)^n T_n\left(\dfrac{x - \sigma/n}{1 + \sigma/n}\right)$ *provided* $0 < \sigma \leqq n\alpha_n$.

Here $u_n(x; A)$ denotes the polynomial (2.4); and in case (a) $A = A(\sigma)$ *is a uniquely determined function which increases monotonically from 1 to $+\infty$ as σ increases from $n\alpha_n$ to $+\infty$; $\alpha_n = \tan^2 [\pi/(2n)]$.*

A corresponding result holds for negative σ, obtained by replacing $f(x)$ by $(-1)^n f(-x)$. For $\sigma = 0$ the extremum is given by Tchebycheff's polynomial.

From (2.4) we obtain, for $x > C$,

$$\begin{aligned} -u_n(x; A) &= Rx^n - Sx^{n-1} + \cdots \\ &= \cosh\left\{n \int_C^x (t - A)(t - B)^{-\frac{1}{2}}(t - C)^{-\frac{1}{2}}(t^2 - 1)^{-\frac{1}{2}}\, dt\right\} \\ &= \cosh\Big\{n\,(\log x - \log C) \\ &\qquad + n \int_C^\infty [(t - A)(t - B)^{-\frac{1}{2}}(t - C)^{-\frac{1}{2}}(t^2 - 1)^{-\frac{1}{2}} - t^{-1}]\, dt \\ &\qquad - n \int_x^\infty [(t - A)(t - B)^{-\frac{1}{2}}(t - C)^{-\frac{1}{2}}(t^2 - 1)^{-\frac{1}{2}} - t^{-1}]\, dt\Big\}; \end{aligned}$$

so that, as $x \to +\infty$,

$$\begin{aligned} -u_n(x; A) = \tfrac{1}{2}(x/C)^n \exp\Big\{&n \int_C^\infty [(t - A)(t - B)^{-\frac{1}{2}}(t - C)^{-\frac{1}{2}}(t^2 - 1)^{-\frac{1}{2}} - t^{-1}]\, dt \\ &- n \int_x^\infty [(\tfrac{1}{2}(B + C) - A)t^{-2} + O(t^{-3})]\, dt\Big\} + O(x^{-n}). \end{aligned}$$

Consequently

$$R = \tfrac{1}{2}C^{-n} \times \exp\left\{n \int_C^\infty [(t - A)(t - B)^{-\frac{1}{2}}(t - C)^{-\frac{1}{2}}(t^2 - 1)^{-\frac{1}{2}} - t^{-1}\, dt\right\} > 0, \tag{8.2}$$

$$S/R = n\{\tfrac{1}{2}(B + C) - A\} > 0,$$

so that R and S are continuous functions of A.

[18] Loc. cit.[3] (a), (b), (c).

From the results of Lemma 1,

$$\left(\frac{d}{dx}\right)^n u_n(x; A) = -n!R, \tag{8.3}$$

is an increasing function of A. Let $A_1 < A_2$, and let R_1, S_1, R_2, S_2 be the corresponding values of R and S, $R_1 > R_2$. Considering $R_1^{-1}u_n(x; A_1) - R_2^{-1}u_n(x; A_2)$ at the extremum points of $u_n(x; A_2)$ in $-1 \leqq x \leqq +1$ we see that it cannot be of degree $n-2$, so that $S_1/R_1 \neq S_2/R_2$. Hence S/R is monotonic. Its minimum value is attained for

$$u_n(x; +1) = -T_n\left(\frac{x-\alpha_n}{1+\alpha_n}\right),$$

so that $\min (S/R) = n\alpha_n$. Its maximum value is attained for $u_n(x; +\infty) = T_{n-1}(x)$, so that $\max (S/R) = +\infty$.

Now let $f(x)$ be a polynomial of the form (8.1), and let $\sigma \geqq n\alpha_n$. Then there exists a definite polynomial $u_n(x; A)$, $A = A(\sigma) \geqq 1$, for which $S/R = \sigma$ so that

$$d(x) = f(x) + R^{-1}u_n(x; A) \tag{8.4}$$

is of degree $n-2$. Let $\max |f(x)| \leqq R^{-1}$, $-1 \leqq x \leqq +1$. Then the polynomial (8.4) is alternately $\geqq 0$ and $\leqq 0$ at the points at which $u_n(x; A) = \pm 1$. Unless $d(x) \equiv 0$ this gives $n-1$ distinct points at which $d'(x)$ is alternately >0 and <0, and hence $n-2$ roots for $d'(x)$ which is impossible.

2. The argument is similar in the other case, $0 < \sigma < n\alpha_n$, since the polynomial

$$2^{1-n}(1+\sigma/n)^n T_n\left(\frac{x-\sigma/n}{1+\sigma/n}\right) = x^n - \sigma x^{n-1} + \cdots \tag{8.5}$$

assumes its maximum modulus $2^{1-n}(1+\sigma/n)^n$ precisely n times in $-1 \leqq x \leqq +1$.

Replacing $-R^{-1}u_n(x; A)$ in (8.4) by the left-hand side of (8.5), we obtain the desired result.

3. Another theorem of Zolotareff is the following[19]:

THEOREM 4. *Let x_0, y_0, be arbitrary real numbers, of which $x_0 > 1$, and let $f(x)$ be an arbitrary polynomial of degree n satisfying the conditions*

$$f(x) = x^n + \cdots, \qquad f(x_0) = y_0. \tag{8.6}$$

Then $\max |f(x)|$, $-1 \leqq x \leqq +1$, *is a minimum if and only if $f(x)$ is one of the polynomials*

$$\begin{gathered} -R^{-1}u_n(x; A), \qquad 2^{1-n}(1+\alpha)^n T_n\left(\frac{x-\alpha}{1+\alpha}\right), \\ 2^{1-n}(1+\alpha)^n T_n\left(\frac{x+\alpha}{1+\alpha}\right), \qquad (-1)^{n-1}R^{-1}u_n(-x; A). \end{gathered} \tag{8.7}$$

[19] Loc. cit.[3] (b), p. 27, (c), p. 371.

Here $A \geqq 1$, $0 \leqq \alpha \leqq \alpha_n = \tan^2 [\pi/(2n)]$ *are certain numbers uniquely determined by* x_0 *and* y_0.

The values of the polynomials (8.7) at $x = x_0$ increase

from $-\infty$ to $2^{1-n}(1+\alpha_n)^n T_n\left(\dfrac{x_0 - \alpha_n}{1+\alpha_n}\right) = \beta$

as A decreases from $+\infty$ to $+1$;

from β to $2^{1-n}T_n(x_0)$ as α decreases from α_n to 0,

from $2^{1-n}T_n(x_0)$ to $2^{1-n}(1+\alpha_n)^n T_n\left(\dfrac{x_0+\alpha_n}{1+\alpha_n}\right) = \beta'$

as α increases from 0 to α_n;

from β' to $+\infty$ as A increases from 1 to $+\infty$,

respectively. These facts determine for a given y_0 the extremum polynomial $f_0(x)$ in question. Indeed, consider the difference $f(x) - f_0(x)$ at the points in $-1 \leqq x \leqq +1$ at which $f_0(x) = \pm 1$, and in addition at $x = x_0$. Since this difference is alternately $\geqq 0$ and $\leqq 0$ at these $n+1$ points, the usual argument gives $n-1$ distinct roots for its derivative [unless $f(x) \equiv f_0(x)$], which is impossible.

4. The problem defined by the condition

$$f^{(k)}(x_0) = y_0 \tag{8.8}$$

where $1 \leqq k \leqq n-1$, $x_0 > 1$, and y_0 is arbitrary, can be treated in a similar manner. For $k = n-1$ we obtain the first problem dealt with above.

9. A further application

The previous considerations furnish another property of the polynomials $u_n(x; A)$ of Zolotareff which play a role in the interesting investigations of W. Markoff [see[4]].

1. We prove the following application of Lemma 1:

THEOREM 5. *Let*

$$1 > x_1 > z_1 > x_2 > z_2 > \cdots > x_{n-2} > z_{n-2} > x_{n-1} > -1 \tag{9.1}$$

be the values of x *characterized by the conditions*

$$u_n(x_\nu ; A) = 0, \qquad \nu = 1, 2, \cdots, n-1, \tag{9.2}$$

$$u_n'(z_\nu ; A) = 0, \qquad \nu = 1, 2, \cdots, n-2; \tag{9.3}$$

then the functions $x_\nu = x_\nu(A)$ *and* $z_\nu = z_\nu(A)$ *increase as* A *increases.*[20]

The roots x_ν of $u_n(x; A)$ satisfy the equation

$$\int_{x_\nu}^{1} (A-t)(B-t)^{-\frac{1}{2}}(C-t)^{-\frac{1}{2}}(1-t^2)^{-\frac{1}{2}}\,dt = (\nu - \tfrac{1}{2})\pi/n, \tag{9.4}$$

$$\nu = 1, 2, \cdots, n-1.$$

[20] Concerning x_ν, see W. Markoff, loc. cit. p. 242. The largest root $D = D(A)$ also increases, as can be concluded from the result of §2, No.[4].

We can assume that $A = A(\rho)$, $B = B(\rho)$, $C = C(\rho)$ are increasing functions of a parameter $\rho, \rho > 0$, all these functions having continuous derivatives. Then $x_\nu = x_\nu(\rho)$ and

$$(A - x_\nu)(B - x_\nu)^{-\frac{1}{2}}(C - x_\nu)^{-\frac{1}{2}}(1 - x_\nu^2)^{-\frac{1}{2}}x_\nu'(\rho)$$

$$= \int_{x_\nu}^{1} \frac{d}{d\rho}\{(A - t)(B - t)^{-\frac{1}{2}}(C - t)^{-\frac{1}{2}}(1 - t^2)^{-\frac{1}{2}}\}\, dt$$

$$= \int_{x_\nu}^{1} (B - t)^{-\frac{1}{2}}(C - t)^{-\frac{1}{2}}(1 - t^2)^{-\frac{1}{2}}\, dt\left\{A'(\rho) - \tfrac{1}{2}B'(\rho)\frac{A - t}{B - t} - \tfrac{1}{2}C'(\rho)\frac{A - t}{C - t}\right\}$$

$$> \left\{A'(\rho) - \tfrac{1}{2}B'(\rho)\frac{A + 1}{B + 1} - \tfrac{1}{2}C'(\rho)\frac{A + 1}{C + 1}\right\}\int_{x_\nu}^{1} (B - t)^{-\frac{1}{2}}(C - t)^{-\frac{1}{2}}(1 - t^2)^{-\frac{1}{2}}\, dt$$

$$> 0$$

since the expression (2.9) increases with ρ.

The assertion about z_ν can be proved in a similar manner.

2. The assertion about z_ν follows also from the following general remark. Suppose the roots of an algebraic equation are real and distinct, and that they are increasing functions of a parameter; then the same holds for the roots of the derivative. Indeed, using the notation above:

$$\frac{1}{z_\nu - x_1} + \frac{1}{z_\nu - x_2} + \cdots + \frac{1}{z_\nu - x_n} = 0, \qquad x_n = D.$$

[Here $x_n = D$ denotes the only root of $u_n(x; A)$ which is > 1.] Differentiating this relation,

$$\sum_{\mu=1}^{n} \frac{z_\nu' - x_\mu'}{(z_\nu - x_\mu)^2} = 0,$$

so that $x_\mu' > 0$ implies $z_\nu' > 0$.

Repeated application of this argument shows that the roots of all derivatives $u_n^{(k)}(x; A)$ increase as A increases.

UNIVERSITY OF PENNSYLVANIA
STANFORD UNIVERSITY

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

ON ABEL'S CONVERSE THEOREM

BY GUIDO FUBINI

(Received March 20, 1942)

1. Introduction

In studying Lie's famous papers[1] on the surfaces of translation, I began to think that it was perhaps useful to consider them as the statement of a theorem which, in Lie's case, is the converse of Abel's theorem, so fundamental in the theory of Abelian integrals. Lie generalized his results to other manifolds of translation. I think it will be possible to look for a generalization in another direction; in other words, it is perhaps possible to find new theorems which may be considered as the converse of Abel's theorem. If we start from a curve of genus p, we obtain some particular manifolds of translation of $p - 1$ dimensions. This has no geometrical meaning if, for instance, $p = 1$; but the question of stating for $p = 1$ a theorem which is the converse of Abel's theorem is always interesting, at least according to my opinion. Also Poincaré occupied himself with Lie's results and looked for some generalizations of the surfaces of translation. Though my point of view is different, there are some analogies between Poincaré's problem[2] and the problem to which this paper is devoted. But the methods of this paper are, I think, quite new; Lie starts from a system of *two partial* differential equations with *one* unknown, whereas I start from *one ordinary differential equation* with *four* unknowns (two of which are functions of a parameter u, and two functions of an independent parameter v). And I succeed in transforming this equation into a *reduced* equation which contains only the *first* derivatives with respect to one of the parameters u, v. After this the problem is easily solved, but there are many particular cases to be considered (see the conclusion of this paper). On the other hand it is not possible to take advantage of Poincaré's methods. We do not suppose from the beginning that the functions which determine our curves are analytic. Their analyticity will be a consequence of the differential equation referred to above. But even if we supposed that they are analytic, it would not be possible to make use of Poincaré's method. Let us suppose that c_1 is a curve defined by analytic functions. I choose on it a point A_1 ; afterward I define another analytic curve c_2 by giving the coordinates of its points A_2 as analytic functions of the slope m of the line A_1A_2 joining A_2 (variable on c_2) with the point A_1 chosen on c_1. These functions may exist in a certain region of the plane of m, which we now consider as a complex variable. If this region does not contain the slope of the line t which is tangent to c_1 at A_1, this tangent will neither intersect c_2, nor be

[1] See *Lie:* Berichte ü. die Verhandl. der Kön. Sächs. Ges. der Wissensch. zu Leipzig, 1896, Bd. 48, p. 141. Comptes Rendus 1892 (1st Sem.), pp. 277, 331.

[2] See Bull. de la Soc. Mathem. de France, 1901, tome 29, p. 61. (See also Journal de Liouville, 1895, tome 1, p. 219.)

tangent to it; and it is possible that all the straight lines of a certain neighborhood of t (for instance, the tangents t' to c_1 at the points A_1' of a certain neighborhood of A_1) have the same property. In this case, and in other analogous cases, it is not possible to make use of Poincaré's methods, and I have not succeeded in demonstrating directly that the preceding suppositions are absurd for the curves which we shall study in our problem. In order to state this problem clearly, I recall that the simplest form of Abel's theorem for the curve of genus 1 is the following. *For a non-rational cubic we can choose the Abelian integral of first kind in such a way that its values at three points of the cubic have a sum equal to zero if and only if the three points are collinear.* (I shall later speak of the generalizations to six points lying on the same conic etc.) I now state the following converse problem. Let us suppose we have three *real* arcs of curves c, γ, C defined by parametric equations in which x, y are Cartesian, or, more generally, non-homogeneous projective coordinates, and u, v, w are the parameters:

$$(c)\begin{cases} x = f_1(u), \\ y = f_2(u), \end{cases} \qquad (\gamma)\begin{cases} x = \varphi_1(v), \\ y = \varphi_2(v), \end{cases} \qquad (C)\begin{cases} x = F_1(w), \\ y = F_2(w), \end{cases}$$

and that three points, one on every curve, are collinear if and only if

$$u + v + w = 0. \tag{1.1}$$

Can we deduce that *the three curves belong to the same curve Γ of third degree and that the corresponding parameters are identical with the Abelian integral of first kind corresponding to this curve*? The answer will be given in this paper; it is positive, but we must take into account the cases where the cubic degenerates (into a conic and a straight line or into three straight lines) or is rational (possesses a double point or a cusp).

We could also state a *more general* problem by substituting

$$\Phi_1(u) + \Phi_2(v) + \Phi_3(w) = 0$$

for (1.1) and by supposing that Φ are continuous differentiable functions of the corresponding parameter; but, by a suitable change of parameters, we realize that the new problem is equivalent to the preceding.

2. The fundamental equation

I choose at random two of the three curves c, γ, C, for instance c, γ, and I suppose that neither the former nor the latter is the line at infinity ($x = \infty$, $y = \infty$). (By means of a collineation we can always satisfy this condition.) On the straight line joining a point (f_1, f_2) of c with a point (φ_1, φ_2) of γ I must find a point (F_1, F_2) of C. Its coordinates will be given by

$$F_1 = kf_1 + (1 - k)\varphi_1, \qquad F_2 = kf_2 + (1 - k)\varphi_2, \tag{1.2}$$

in which k is unknown. This point belongs to C; therefore its coordinates must be functions of $w = -(u + v)$, (see (1.1)). Therefore

$$\frac{\partial F_1}{\partial u} = \frac{\partial F_1}{\partial v}, \qquad \frac{\partial F_2}{\partial u} = \frac{\partial F_2}{\partial v},$$

or

$$\left[\frac{\partial k}{\partial u} - \frac{\partial k}{\partial v}\right][f_i(u) - \varphi_i(v)] + k[f_i'(u) + \varphi_i'(v)] = \varphi_i'(v),$$

$$(i = 1, 2) \qquad \left[f' = \frac{dk}{du}, \varphi' = \frac{d\varphi}{dv}\right].$$

We have obtained two linear equations in two unknowns: k and $\partial k/\partial u - \partial k/\partial v$. By solving these equations, we obtain

$$\Delta k = \varphi_2'[f_1 - \varphi_1] - \varphi_1'[f_2 - \varphi_2], \tag{2.2}$$

$$\Delta\left(\frac{\partial k}{\partial u} - \frac{\partial k}{\partial v}\right) = \varphi_1'(f_2' + \varphi_2') - \varphi_2'(f_1' + \varphi_1') = \varphi_1'f_2' - \varphi_2'f_1' \tag{3.2}$$

in which

$$\Delta = (f_1 - \varphi_1)(f_2' + \varphi_2') - (f_2 - \varphi_2)(f_1' + \varphi_1'). \tag{4.2}$$

By differentiating (2.2) with respect to u or to v, and by subtracting, we obtain:

$$\left(\frac{\partial \Delta}{\partial u} - \frac{\partial \Delta}{\partial v}\right)k + \Delta\left(\frac{\partial k}{\partial u} - \frac{\partial k}{\partial v}\right) = \varphi_2'f_1' - \varphi_1'f_2' + \varphi_1''(f_2 - \varphi_2) - \varphi_2''(f_1 - \varphi_1),$$

$$\left(f'' = \frac{d^2f}{du^2}; \varphi'' = \frac{d^2\varphi}{dv^2}\right).$$

By recalling (3.2) we deduce:

$$\left(\frac{\partial \Delta}{\partial u} - \frac{\partial \Delta}{\partial v}\right)k = 2(\varphi_2'f_1' - \varphi_1'f_2') + \varphi_1''(f_2 - \varphi_2) - \varphi_2''(f_1 - \varphi_1).$$

By multiplying by Δ, by comparing with (2.2), and by remarking that

$$\frac{\partial \Delta}{\partial u} - \frac{\partial \Delta}{\partial v} = (f_2'' - \varphi_2'')(f_1 - \varphi_1) - (f_2 - \varphi_2)(f_1'' - \varphi_1'')$$

we conclude that

$$\begin{aligned}&\{(f_2'' - \varphi_2'')(f_1 - \varphi_1) - (f_2 - \varphi_2)(f_1'' - \varphi_1'')\}\{\varphi_2'(f_1 - \varphi_1) - \varphi_1'(f_2 - \varphi_2)\}\\ &= \{(f_1 - \varphi_1)(f_2' + \varphi_2') - (f_2 - \varphi_2)(f_1' + \varphi_1')\}\\ &\qquad \{2(\varphi_2'f_1' - \varphi_1'f_2') + \varphi_1''(f_2 - \varphi_2) - \varphi_2''(f_1 - \varphi_1)\},\end{aligned} \tag{5.2}$$

or

$$\begin{aligned}&\{\varphi_2'(f_1 - \varphi_1) - \varphi_1'(f_2 - \varphi_2)\}\{f_2''(f_1 - \varphi_1) - f_1''(f_2 - \varphi_2) - 2(\varphi_2'f_1' - \varphi_1'f_2')\}\\ &= \{f_2'(f_1 - \varphi_1) - f_1'(f_2 - \varphi_2)\}\\ &\qquad \{\varphi_1''(f_2 - \varphi_2) - \varphi_2''(f_1 - \varphi_1) + 2(\varphi_2'f_1' - \varphi_1'f_2')\}.\end{aligned} \tag{6.2}$$

The second member can be deduced from the first by interchanging the f, φ and changing the sign. This is the *fundamental equation* of our problem. If this equation is satisfied, and if $\Delta \neq 0$, the (2.2) gives us a value of k such that the functions F_1, F_2 determined by (1.2) are functions of $w = -(u + v)$ and the three curves c, γ, C satisfy the conditions of our problem. If $\Delta = 0$, the same method which led us to (5.2) or (6.2) proves that these equations are satisfied, which can be checked directly.

If $\Delta = 0$, we have

$$(f_1 - \varphi_1)\varphi_2' - (f_2 - \varphi_2)\varphi_1' = -\{(f_1 - \varphi_1)f_2' - (f_2 - \varphi_2)f_1'\},$$

so that (5.2) or (6.2) are equivalent to the equation

$$\{\varphi_2'(f_1 - \varphi_1) - \varphi_1'(f_2 - \varphi_2)\}\{(f_2'' - \varphi_2'')(f_1 - \varphi_1) - (f_2 - \varphi_2)(f_1'' - \varphi_1'')\} = 0$$

or

$$\{\varphi_2'(f_1 - \varphi_1) - \varphi_1'(f_2 - \varphi_2)\}\left(\frac{\partial \Delta}{\partial u} - \frac{\partial \Delta}{\partial v}\right) = 0,$$

which is a consequence of the equation $\Delta = 0$. Consequently the fundamental equations are satisfied if $\Delta = 0$; but, in this case, the (2.2) does not determine the value of k. But, if we neglect the case withont any geometrical meaning (at least for our problem) that c and γ are *pieces of the same straight line*, the slope

$$m = \frac{f_2 - \varphi_2}{f_1 - \varphi_1}$$

of the line joining the points (f_1, f_2) of c and (φ_1, φ_2) of γ cannot be a constant. And the equation $\Delta = 0$ is equivalent to the equation

$$\frac{\partial m}{\partial u} = \frac{\partial m}{\partial v},$$

which proves that m is a function of $w = -(u + v)$. In the exceptional case $\Delta = 0$, the conditions of our problem are satisfied too; but the third line C is the line at infinity of the plane (x, y). Therefore (if neither c, nor γ is the line at infinity) *the fundamental equation* gives us all the necessary and sufficient conditions for these lines c, γ.

3. Remarks on the fundamental equation

We can arrive at this equation also by another way. If three points (f_1, f_2) of c, (φ_1, φ_2) of γ and (F_1, F_2) are collinear, we can suppose that $y = mx + n$ is the equation of the line joining them. Therefore

(1.3) $f_2 = mf_1 + n$, (2.3) $\varphi_2 = m\varphi_1 + n$, (3.3) $F_2 = mF_1 + n$,

from which we deduce

(4.3) $$m = \frac{f_2 - \varphi_2}{f_1 - \varphi_1}.$$

By differentiating (1.3) with respect to v, and (2.3) by respect to u, we deduce that

$$\frac{\partial n}{\partial v} = -f_1 \frac{\partial m}{\partial v}, \qquad \frac{\partial n}{\partial u} = -\varphi_1 \frac{\partial m}{\partial u}. \tag{5.3}$$

Since F_i are functions of $w = -(u + v)$, we have (if $i = 1$, or 2):

$$\frac{\partial F_i}{\partial u} = \frac{\partial F_i}{\partial v}.$$

From (3.3) and (5.3) we deduce consequently

$$\frac{\partial m}{\partial u}(F_1 - \varphi_1) = \frac{\partial m}{\partial v}(F_1 - f_1), \quad \text{or} \quad F_1 = \frac{\varphi_1 \dfrac{\partial m}{\partial u} - f_1 \dfrac{\partial m}{\partial v}}{\dfrac{\partial m}{\partial u} - \dfrac{\partial m}{\partial v}}. \tag{6.3}$$

By comparing with (1.2) we deduce that

$$k = -\frac{\dfrac{\partial m}{\partial v}}{\dfrac{\partial m}{\partial u} - \dfrac{\partial m}{\partial v}},$$

and this equation is identical with (2.2). The fundamental equation (6.2) states only that F_1 defined by (6.3) is a function of $w = -(u + v)$.

The same method which led us to the fundamental equation proves that this equation is invariant under the group of collineations in the plane (x, y). In order to check this statement, we shall make use of *homogeneous* projective coordinates, by writing

$$f_1 = x = \frac{x_1}{x_3}, \qquad f_2 = y = \frac{x_2}{x_3}, \qquad \varphi_1 = \frac{\xi_1}{\xi_3}, \qquad \varphi_2 = \frac{\xi_2}{\xi_3}.$$

I shall indicate by

$$(x\xi'\xi) \text{ or } (x''\xi x), \text{ etc.},$$

the determinants of third order, the r^{th} row of which ($r = 1, 2, 3$) is

$$x_r\,,\ \xi'_r\,,\ \xi_r \text{ or } x''_r\,,\ \xi_r\,,\ x_r\,, \text{ etc.}$$

The fundamental equation can be written in the form

$$\begin{vmatrix} f_1 - \varphi_1 & \varphi'_1 \\ f_2 - \varphi_2 & \varphi'_2 \end{vmatrix} \begin{vmatrix} f_1 - \varphi_1 & f''_1 \\ f_2 - \varphi_2 & f''_2 \end{vmatrix} + \begin{vmatrix} f_1 - \varphi_1 & f'_1 \\ f_2 - \varphi_2 & f'_2 \end{vmatrix} \begin{vmatrix} f_1 - \varphi_1 & \varphi''_1 \\ f_2 - \varphi_2 & \varphi''_2 \end{vmatrix}$$

$$= 2 \begin{vmatrix} f'_1 & \varphi'_1 \\ f'_2 & \varphi'_2 \end{vmatrix} \begin{vmatrix} f_1 - \varphi_1 & f'_1 + \varphi'_1 \\ f_2 - \varphi_2 & f'_2 + \varphi'_2 \end{vmatrix},$$

or, by supposing at first $x_3 = \xi_3 = 1$ and consequently $x_i = f_i$, $\xi_i = \varphi_i$ $(i = 1, 2)$

$$(x\xi'\xi)(xx''\xi) + (xx'\xi)(x\xi''\xi) = 2(x'x\xi)\begin{vmatrix} f_1' & \varphi_1' & A \\ f_2' & \varphi_2' & B \\ 0 & 0 & 1 \end{vmatrix} + 2(x\xi'\xi)\begin{vmatrix} f_1' & \varphi_1' & C \\ f_2' & \varphi_2' & D \\ 0 & 0 & 1 \end{vmatrix},$$

in which A, B, C, D are arbitrary. By supposing $A = \varphi_1$, $B = \varphi_2$, $C = f_1$, $D = f_2$, we find

$$(7.3) \qquad (x\xi'\xi)(xx''\xi) + 2(x'\xi'\xi)(\xi x'x) = (\xi''x\xi)(xx'\xi) + 2(x\xi'\xi)(x'\xi'x).$$

But it is very easy to see that if we substitute ρx_i and $\sigma\xi_i$ for x_i, ξ_i, and by supposing that ρ is a function of the only u, and σ a function of the only v, this equation does not change, because both of its members are multiplied by the same factor $\rho^3\sigma^3$. It is consequently useless to suppose $x_3 = \xi_3 = 1$. They can be arbitrary functions of u, v, respectively.

Moreover, if I transform the x, ξ by means of the same linear homogeneous transformation,[3] both the members of (7.3) are multiplied by the *same* factor (the square of the determinant of the transformation). Therefore the form (7.3) of the fundamental equation proves that it is invariant under the group of the collineation. The two members divided, for instance, by

$$(xx'x'')(\xi\xi'\xi'')$$

are *absolute projective invariants* of the pair of curves c, γ; they have a geometrical meaning if the parameters u, v have a geometrical meaning (for instance, if they are the projective length of an arc of the curves). I think that nobody has noticed these invariants heretofore.

4. Reduction to the first order

It does not seem to be an easy matter to study directly the fundamental equation; but we can succeed in transforming it into an equation in which there are only derivatives of first order, for instance with respect to u. In the following lines, if Φ is a function of u, v, I shall indicate by Φ' the derivative $\partial\Phi/\partial u$. For short, I shall write

$$t_i = f_i - \varphi_i, \qquad t_i' = f_i', \qquad t_i'' = f_i'' \qquad (i = 1, 2).$$

The fundamental equation can be written in the form

$$(\varphi_2't_1 - \varphi_1't_2)(t_2''t_1 - t_1''t_2) - 2(\varphi_2't_1' - \varphi_1't_2')(t_2't_1 - t_2t_1')$$
$$= 2(\varphi_2't_1' - \varphi_1't_2')(\varphi_2't_1 - \varphi_1't_2) + (t_2't_1 - t_1't_2)(\varphi_1''t_2 - \varphi_2''t_1).$$

Dividing both members by $(\varphi_2't_1 - \varphi_1't_2)^3$ I obtain

$$\frac{d}{dt}\frac{t_2't_1 - t_1't_2}{(\varphi_2't_1 - \varphi_1't_2)^2} = 2\frac{\varphi_2't_1' - \varphi_1't_2'}{(\varphi_2't_1 - \varphi_1't_2)^2} + \frac{t_2't_1 - t_2t_1'}{(\varphi_2't_1 - \varphi_1't_2)^2}\frac{\varphi_1''t_2 - \varphi_2''t_1}{\varphi_2't_1 - \varphi_1't_2}.$$

[3] It is now sufficient to suppose that the coefficients of the transformation are constant.

Let A, B be any two functions of v only. It is obvious that

$$\frac{d}{du}\frac{At_1 - Bt_2}{\varphi_2' t_1 - \varphi_1' t_2} = (A\varphi_1' - B\varphi_2')\frac{t_1 t_2' - t_2 t_1'}{(\varphi_2' t_1 - \varphi_1' t_2)^2}$$

$$\frac{\varphi_1'' t_2 - \varphi_2'' t_1}{\varphi_2' t_1 - \varphi_1' t_2} = \frac{\varphi_1'' \varphi_2' - \varphi_1' \varphi_2''}{A\varphi_1' - B\varphi_2'}\frac{At_1 - Bt_2}{\varphi_2' t_1 - \varphi_1' t_2} + \frac{B\varphi_2'' - A\varphi_1''}{A\varphi_1' - B\varphi_2'},$$

so that our equation becomes

$$\frac{d^2}{du^2}\frac{At_1 - Bt_2}{\varphi_2' t_1 - \varphi_1' t_2} = -2(A\varphi_1' - B\varphi_2')\frac{d}{du}\frac{1}{\varphi_2' t_1 - \varphi_1' t_2}$$

$$+\left(\frac{\varphi_1'' \varphi_2' - \varphi_1' \varphi_2''}{A\varphi_1' - B\varphi_2'}\frac{At_1 - Bt_2}{\varphi_2' t_1 - \varphi_1' t_2} + \frac{B\varphi_2'' - A\varphi_1''}{A\varphi_1' - B\varphi_2'}\right)\frac{d}{du}\frac{At_1 - Bt_2}{\varphi_2' t_1 - \varphi_1' t_2}.$$

By integrating with respect to u, we deduce

$$(1.4)\quad \begin{cases} \dfrac{d}{du}\dfrac{At_1 - Bt_2}{\varphi_2' t_1 - \varphi_1' t_2} = -2(A\varphi_1' - B\varphi_2')\dfrac{1}{\varphi_2' t_1 - \varphi_1' t_2} \\ \qquad + \dfrac{B\varphi_2'' - A\varphi_1''}{A\varphi_1' - B\varphi_2'}\dfrac{At_1 - Bt_2}{\varphi_2' t_1 - \varphi_1' t_2} + \dfrac{1}{2}\dfrac{\varphi_1'' \varphi_2' - \varphi_1' \varphi_2''}{A\varphi_1' - B\varphi_2'}\left(\dfrac{At_1 - Bt_2}{\varphi_2' t_1 - \varphi_1' t_2}\right)^2 + D, \end{cases}$$

in which D is a new (unknown) function of v, whereas A, B are functions of v, arbitrarily chosen, such that

$$(2.4)\qquad A\varphi_1' - B\varphi_2' \neq 0.$$

(If we change A and B, the function D also will change.)

We will give to (1.4) the name of *reduced fundamental equation*. This reduction allows us to solve our problem. If v is a point of γ at which $\varphi_1' = 0$, I may suppose $B = 1$, because we can always suppose that we consider points at which the direction cosines of the tangent to γ are determined, by stating that they are proportional to φ_1', φ_2'; so that, if $\varphi_1' = 0$, we conclude that $\varphi_2' \neq 0$; and (2.4) is satisfied if $B = 1$.

We may also suppose $A = 0$; it is easy to check that a change of A corresponds only to a change of D, and therefore we conclude:

$$(3.4)\quad \frac{d}{du}\left(\frac{f_2 - \varphi_2}{f_1 - \varphi_1}\right) = \frac{d}{du}\left(\frac{t_2}{t_1}\right) = -2\varphi_2'\frac{1}{f_1 - \varphi_1} - \frac{\varphi_2'' f_2 - \varphi_2}{\varphi_2' f_1 - \varphi_1}$$

$$+\frac{\varphi_1''}{2\varphi_2'}\left(\frac{f_2 - \varphi_2}{f_1 - \varphi_1}\right)^2 + D_1.$$

In this equation v is no longer a variable, because we have considered a *particular* point of γ at which $\varphi_1' = 0$; therefore φ_2', φ_2'', φ_1'', D_1 are constants in (3.4).

We could consider v variable in (3.4) only if $\varphi_1' = 0$ is an identity, or if $x = \varphi_1 = \text{const.}$ is the equation of γ. In the same manner, we find that

$$(3.4)\text{ bis}\qquad \frac{d}{du}\left(\frac{f_1-\varphi_1}{f_2-\varphi_2}\right) = -2\varphi_1'\frac{1}{f_2-\varphi_2} - \frac{\varphi_1''}{\varphi_1'}\frac{f_1-\varphi_1}{f_2-\varphi_2} + \frac{\varphi_2''}{2\varphi_1'}\left(\frac{f_1-\varphi_1}{f_2-\varphi_2}\right)^2 + D_2$$

(if $\varphi_2' = 0$). In the general case $\varphi_1'\varphi_2' \neq 0$, the equation (1.4) may be written in the form:

$$t_1t_2' - t_2t_1' = -2(\varphi_2't_1 - \varphi_1't_2) + \frac{1}{(A\varphi_1' - B\varphi_2')^2}\left\{(B\varphi_2'' - A\varphi_1'')(At_1 - Bt_2)(\varphi_2't_1 - \varphi_1't_2) + \frac{(\varphi_1''\varphi_2' - \varphi_1'\varphi_2'')(At_1 - Bt_2)^2}{2}\right\} + \frac{D}{A\varphi_1' - B\varphi_2'}(\varphi_2't_1 - \varphi_1't_2)^2.$$

The second member must obviously be independent of A, B (but D may change if we change the functions A, B). In order to check this statement it is sufficient to remark that the equation

$$2(B\varphi_2'' - A\varphi_1'')(At_1 - Bt_2)(\varphi_2't_1 - \varphi_1't_2) + (\varphi_1''\varphi_2' - \varphi_1'\varphi_2'')(At_1 - Bt_2)^2 = -(lA^2 + 2mAB + nB^2)(\varphi_2't_1 - \varphi_1't_2)^2 + (nt_1^2 + 2mt_1t_2 + lt_2^2)(\varphi_2'B - \varphi_1'A)^2$$

is an *identity* if l, m, n are functions of v satisfying the two following linear equations:

$$(4.4)\qquad l\varphi_2'^2 - n\varphi_1'^2 = \varphi_1''\varphi_2' + \varphi_2''\varphi_1'; \qquad m\varphi_1' = \varphi_1'' - l\varphi_2'$$
$$(\text{or } m\varphi_2' = -\varphi_2'' - n\varphi_1').$$

Our equation becomes consequently

$$(t_1t_2' - t_2t_1') = -2(\varphi_2't_1 - \varphi_1't_2) + \tfrac{1}{2}(nt_1^2 + 2mt_1t_2 + lt_2^2) + \left[\frac{D}{A\varphi_1' - B\varphi_2'} - \frac{lA^2 + 2mAB + nB^2}{2(A\varphi_1' - B\varphi_2')^2}\right](\varphi_2't_1 - \varphi_1't_2)^2.$$

But I may suppose that l, m, n satisfy not only (4.4) but also

$$(5.4)\qquad 2D(A\varphi_1' - B\varphi_2') = lA^2 + 2mAB + nB^2.$$

The equations (4.4) and (5.4) are a system of three linear equations in l, m, n; and it is easy to see that one can solve these equations with respect to l, m, n if $\varphi_1'\varphi_2' \neq 0$ and $A\varphi_1' - B\varphi_2' \neq 0$. We obtain

$$t_1t_2' - t_2t_1' = -2(\varphi_2't_1 - \varphi_1't_2) + \tfrac{1}{2}(nt_1^2 + 2mt_1t_2 + lt_2^2)$$

in which l, m, n are functions of v satisfying (4.4). (We can deduce nothing from (5.4), because D is unknown.) By writing

$$l\varphi_2'^2 + n\varphi_1'^2 = 4H\varphi_1'\varphi_2'$$

we deduce:

$$\tfrac{1}{2}l\varphi_2'^2 = \tfrac{1}{4}(\varphi_1''\varphi_2' + \varphi_2''\varphi_1') + H\varphi_1'\varphi_2', \qquad \tfrac{1}{2}n\varphi_1'^2 = -\tfrac{1}{4}(\varphi_1''\varphi_2' + \varphi_1'\varphi_2'') + H\varphi_1'\varphi_2',$$
$$\tfrac{1}{2}m\varphi_1'\varphi_2' = \tfrac{1}{2}\varphi_1''\varphi_2' - \tfrac{1}{2}l\varphi_2'^2 = \tfrac{1}{4}(\varphi_1''\varphi_2' - \varphi_1'\varphi_2'') - H\varphi_1'\varphi_2',$$

so that our equation becomes

$$t_1t_2' - t_2t_1' = -2(\varphi_2't_1 - \varphi_1't_2) + H\left(\frac{t_1}{\varphi_1'} - \frac{t_2}{\varphi_2'}\right)^2\varphi_1'\varphi_2'$$
$$+ \tfrac{1}{4}(\varphi_1''\varphi_2' + \varphi_2''\varphi_1')\left[\left(\frac{t_2}{\varphi_2'}\right)^2 - \left(\frac{t_1}{\varphi_1'}\right)^2\right] + 2\tfrac{1}{4}(\varphi_1''\varphi_2' - \varphi_1'\varphi_2'')\frac{t_1t_2}{\varphi_1'\varphi_2'}.$$

If we divide this equation by $\varphi_1'\varphi_2'$ and put

$$\frac{f_1 - \varphi_1}{\varphi_1'} = z_1, \qquad \frac{f_2 - \varphi_2}{\varphi_2'} = z_2 \tag{6.4}$$

we obtain

$$z_1z_2' - z_2z_1' = 2(z_2 - z_1) + H(z_1 - z_2)^2 + \tfrac{1}{4}\left(\frac{\varphi_1''}{\varphi_1'} + \frac{\varphi_2''}{\varphi_2'}\right)(z_2^2 - z_1^2)$$
$$+ \tfrac{1}{2}\left(\frac{\varphi_1''}{\varphi_1'} - \frac{\varphi_2''}{\varphi_2'}\right)z_1z_2. \tag{7.4}$$

This equation is equivalent to the reduced fundamental equation (if we interchange the indices 1, 2, we have to change only the sign of H). If we put

$$H = D\varphi_1'\varphi_2' - \frac{1}{4}\left(\frac{\varphi_1''}{\varphi_1'} + \frac{\varphi_2''}{\varphi_2'}\right)$$

this equation becomes

$$(f_1 - \varphi_1)f_2' - (f_2 - \varphi_2)f_1' = -2[\varphi_2'(f_1 - \varphi_1) - \varphi_1'(f_2 - \varphi_2)]$$
$$- \frac{1}{2}\frac{\varphi_1''\varphi_2' + \varphi_2''\varphi_1'}{(\varphi_1')^2}(f_1 - \varphi_1)^2 \tag{8.4}$$
$$+ \frac{\varphi_1''}{\varphi_1'}(f_1 - \varphi_1)(f_2 - \varphi_2) + D\{\varphi_2'(f_1 - \varphi_1) - \varphi_1'(f_2 - \varphi_2)\}^2,$$

or

$$\frac{d}{du}\frac{f_1 - \varphi_1}{\Omega} = -\frac{2}{\Omega}\varphi_1' + \frac{\varphi_1''\varphi_2' - \varphi_1'\varphi_2''}{2\varphi_1'}\left(\frac{f_1 - \varphi_1}{\Omega}\right)^2 - \frac{\varphi_1''}{\varphi_1'}\frac{f_1 - \varphi_1}{\Omega} + D\varphi_1' \tag{9.4}$$
$$[\Omega = \varphi_2'(f_1 - \varphi_1) - \varphi_1'(f_2 - \varphi_2)].$$

5. Remarks on the reduced fundamental equation

We could repeat our considerations by substituting the curve C for the curve c. In this manner we arrive at another equation which may be deduced

by (7.4) or (6.4) by substituting F_1, F_2, w for f_1, f_2, u, and by changing, if necessary, the function H of the function D. We can demonstrate that it is sufficient to substitute F_1, F_2, w for f_1, f_2, u, *without any change of the functions H or D*. Since $w = -(u + v)$, we deduce that

$$\frac{d}{dw} = -\frac{d}{du}$$

(if v is considered as a constant). If we write the equations

$$Z_1 = \frac{F_1 - \varphi_1}{\varphi_1'}, \qquad Z_2 = \frac{F_2 - \varphi_2}{\varphi_2'}$$

[which are analogous to (6.4)], we deduce from (1.2) that

$$Z_1 = kz_1, \qquad Z_2 = kz_2, \qquad Z' = \frac{\partial Z}{\partial w} = -\frac{\partial}{\partial u}(kz),$$

and therefore

$$Z_1\frac{\partial Z_2}{\partial w} - Z_2\frac{\partial Z_1}{\partial w} = -k^2\left(z_1\frac{\partial z_2}{\partial u} - z_2\frac{\partial z_1}{\partial u}\right) = -k^2(z_1z_2' - z_2z_1').$$

The equation, deduced from (7.4) by substituting Z, w for z, u, will be

$$-(z_1z_2' - z_2z_1') = \frac{2}{k}(z_2 - z_1) + \bar{H}(z_1 - z_2)^2 + \frac{1}{4}\left(\frac{\varphi_1''}{\varphi_1'} + \frac{\varphi_2''}{\varphi_2'}\right)(z_2^2 - z_1^2) + \frac{1}{2}\left(\frac{\varphi_1''}{\varphi_1'} - \frac{\varphi_2''}{\varphi_2'}\right)z_1z_2$$

in which $\bar{H}$ plays, for the curve C, the same role played by H for the curve c. In order to prove that $\bar{H} = H$, it is consequently necessary to prove that

$$-(z_1z_2' - z_2z_1') - \frac{2}{k}(z_2 - z_1) = z_1z_2' - z_2z_1' - 2(z_2 - z_1),$$

or

$$z_2z_1' - z_1z_2' + (z_2 - z_1) = \frac{1}{k}(z_2 - z_1),$$

and this is an obvious consequence of (2.2).

The reduced equation for f_1, f_2 depends on a parameter v. In *the most general* case we can *suppose* it is possible to give to v three different values such that the *three* equations deduced in this way from the reduced equation are independent of each other (none of them is a consequence of the other two. (Obviously this supposition *has to be demonstrated.*) If it is so, by eliminating the *two* derivatives f_1', f_2', we obviously obtain an *algebraic equation* between f_1, f_2; and therefore the curve c would be algebraic. But the curves c and C satisfy the same reduced equation, as we have just proved. Therefore C and c would belong to

the same algebraic curve. And since c and C are any two curves chosen from among the three curves c, C, γ we can conclude that, *some particular cases excepted, the three curves* c, C, γ, *are arcs of the same algebraic curves.*[4]

But it is necessary to study the question more deeply in order to find the exceptional cases and to prove that the quoted algebraic curve is a cubic.

Other remarks on the reduced fundamental equation.

If we apply a collineation, the reduced fundamental equation has to be transformed into another equation which may be deduced from the initial one by changing, if necessary, the function D or H. We shall check this statement in a case of great importance for us. Instead of writing that the coordinates x, y (of a point of c) and the coordinates ξ, η (of a point of γ) are equal to f_1, f_2, φ_1, φ_2 respectively, we suppose that

$$x = \frac{f_1}{f_2}, \quad y = \frac{1}{f_2}, \quad f_1 = \frac{x}{y}, \quad f_2 = \frac{1}{y}, \quad \xi = \frac{\varphi_1}{\varphi_2}, \quad \eta = \frac{1}{\varphi_2}, \quad \varphi_1 = \frac{\xi}{\eta}, \quad \varphi_2 = \frac{1}{\eta}.$$

We shall write also

$$\omega = \eta'(x = \xi) - \xi'(y - \eta),$$

and we find, without difficulty, that

$$\frac{f_1 - \varphi_1}{\Omega} = -\eta\left(\frac{\xi'\eta - \xi\eta'}{\xi'}\frac{x - \xi}{\omega} + \frac{\xi}{\xi'}\right),$$

$$\frac{d}{du}\left(\frac{f_1 - \varphi_1}{\Omega}\right) = \eta(\xi'\eta - \xi\eta')\frac{x'(y - \eta) - y'(x - \xi)}{\omega^2},$$

$$\frac{\varphi_1''\varphi_2' - \varphi_1'\varphi_2''}{2\varphi_1'} = \frac{1}{2\eta}\frac{\xi'\eta'' - \eta'\xi''}{\xi'\eta - \xi\eta'}; \qquad \frac{\varphi_1''}{\varphi_1'} = \frac{\xi''\eta - \xi\eta''}{\xi'\eta - \xi\eta'} - 2\frac{\eta'}{\eta},$$

$$-2\frac{\varphi_1'}{\Omega} = 2y\frac{\xi'\eta - \xi\eta'}{\omega} - (\xi'\eta - \xi\eta')\left[\frac{2\eta}{\omega} - \frac{2}{\xi'} + 2\frac{\eta'}{\xi'}\frac{x - \xi}{\omega}\right].$$

Our equation (9.4) becomes:

$$\eta(\xi'\eta - \xi\eta')\frac{x'(y - \eta) - y'(x - \xi)}{\omega^2} = 2(\xi'\eta - \xi\eta')\left(\frac{\eta}{\omega} + \frac{\eta'}{\xi'}\frac{x - \xi}{\omega}\right)$$
$$+ \frac{\eta}{2}\frac{\xi'\eta'' - \eta'\xi''}{\xi'\eta - \xi\eta'}\left\{\left(\frac{\xi'\eta - \xi\eta'}{\xi'}\frac{x - \xi}{\omega}\right)^2 + 2\frac{\xi(\xi'\eta - \xi\eta')}{\xi'^2}\frac{x - \xi}{\omega}\right\}$$
$$+ \left(\frac{\xi''\eta - \xi\eta''}{\xi'\eta - \xi\eta'} - 2\frac{\eta'}{\eta}\right)\eta\frac{\xi'\eta - \xi\eta'}{\xi'}\frac{x - \xi}{\omega} + \Delta\eta(\xi'\eta - \xi\eta')$$

[4] We can arrive at the same results by starting from the initial (non-reduced) fundamental equation (6.2). We ought to suppose that we can choose *five* values of v such that the corresponding *five* equations are independent. Therefore we could eliminate the *four* derivatives f_1', f_2', f_1'', f_2'' and obtain an *algebraic* equation between f_1, f_2. This equation would also be satisfied by F_1, F_2. But, when one follows this course, the research is much more difficult and long.

(in which Δ depends on v only), or

$$\frac{x'(y-\eta)-y'(x-\xi)}{\omega^2}=\frac{2}{\omega}+\frac{\xi'\eta''-\eta'\xi''}{2\xi'^2}\left(\frac{x-\xi}{\omega}\right)^2+\frac{\xi''}{\xi'^2}\frac{x-\xi}{\omega}+\Delta(v),$$

or

$$\frac{d}{du}\left(\frac{x-\xi}{\omega}\right)=-2\frac{\xi'}{\omega}+\frac{\xi''\eta'-\xi'\eta''}{2\xi'}\left(\frac{x-\xi}{\omega}\right)^2-\frac{\xi''}{\xi'}\frac{x-\xi}{\omega}-\Delta(v)\xi',$$

which can be deduced from (9.4) by changing the value of D and by substituting x, y, ξ, η, ω for $f_1, f_2, \varphi_1, \varphi_2, \Omega$ respectively. Moreover ω can be deduced from Ω by substituting x, y, ξ, η for $f_1, f_2, \varphi_1, \varphi_2$ respectively. We have consequently checked that (9.4) is invariant with respect to the collineation which we have considered.

6. Two of the three curves c, γ, C are straight lines

In this case I may suppose that c, γ are straight lines and suppose moreover (by using a collineation) that they are the lines $y = 0$ and $x = 0$, so that $f_2 = \varphi_1 = 0$. The fundamental equation becomes

$$\frac{f_1''}{f_1'}-2\frac{f_1'}{f_1}=-\left(\frac{\varphi_2''}{\varphi_2'}-2\frac{\varphi_2'}{\varphi_2}\right).$$

Since the first member is independent of v, and the second of u, both of them are equal to the same constant h, so that

$$\frac{d}{du}\log\frac{f_1'}{f_1^2}=h=-\frac{d}{dv}\log\frac{\varphi_2'}{\varphi_2^2}. \tag{1.6}$$

If

$$\Delta=f_1\varphi_2'+\varphi_2 f_1'$$

should be equal to zero, we already know that the third line C would be the line at infinity. We have therefore only to study our problem if $\Delta \neq 0$. (Moreover, by means of a collineation, we could always exclude the case that C is the line at infinity.) From (1.2) and (2.2) we deduce that

$$X=F_1=kf_1=\frac{1}{\frac{1}{f_1}+\frac{\varphi_2 f_1'}{\varphi_2' f_1^2}} \quad\text{and}\quad Y=\frac{1}{\frac{1}{\varphi_2}+\frac{\varphi_2' f_1}{\varphi_2^2 f_1'}} \tag{2.6}$$

are the parametric equations of the third curve C. If $h = 0$ we deduce from (1.6) that

$$\frac{1}{f_1}=Au+B,\qquad \frac{1}{\varphi_2}=Bv+M,\qquad (A, L, B, M = \text{const.})$$

$$X=\frac{1}{A(u+v)+\frac{MA+LB}{B}},\qquad Y=\frac{1}{B(u+v)+\frac{MA+LB}{A}}, \tag{3.6}$$

which are functions of $w = -(u + v)$. The third line has the equation $BY - AX = 0$ and is therefore a third straight line belonging to the pencil determined by the lines c, γ. In this case the coordinates of the pencil of every one of our curves are *rational* functions of the corresponding parameters. The three curves c, γ, C form together a cubic with a triple point (which degenerates consequently into three straight lines).

If $h \neq 0$, we obtain analogously from (1.6) and (2.6),

$$\frac{1}{f_1} = Ae^{hu} + M, \qquad \frac{1}{\varphi_2} = Be^{-hv} + M,$$

$$X = \frac{1}{M - \dfrac{NA}{B}e^{h(u+v)}}, \qquad Y = \frac{1}{N - \dfrac{MB}{A}e^{-h(u+v)}} \tag{4.6}$$

(A, B, M, N = const.) The third curve satisfies the linear equation $MX + NY = 1$, and is again a straight line which *does not belong* to the pencil determined by c, γ; the coordinates of the points of every one of our lines are *not rational* functions of the corresponding parameters: it is necessary to make use of the *exponential* function. (The same thing is true if $\Delta = 0$.) The three lines c, γ, C form *a cubic with three double points*. In any case, if *two of our curves are straight lines, the third line is also a straight line.*

7. Only one of the three curves c, γ, C is a straight line

We suppose that γ is a straight line so that, by means of a collineation I may suppose $\varphi_2 = 0$ (identically). The corresponding *reduced* fundamental equation (3.4) becomes (if we write φ for φ_1):

$$\frac{d}{du}\left(\frac{f_1 - \varphi}{f_2}\right) = -2\varphi'\frac{1}{f_2} - \frac{\varphi''}{\varphi'}\frac{f_1 - \varphi}{f_2} + \Phi(v)$$

[Φ a function of v only] or

$$\left(\frac{f_1}{f_2}\right)' - \varphi\left(\frac{1}{f_2}\right)' = \frac{\varphi\varphi'' - 2\varphi'^2}{\varphi'}\frac{1}{f_2} - \frac{\varphi''}{\varphi'}\frac{f_1}{f_2} + \Phi(v). \tag{1.7}$$

By giving to v two values v_1, v_2 such that $\varphi(v_1) \neq \varphi(v_2)$, we obtain two linear equations in $(f_1/f_2)'$ and $(1/f_2)'$. By solving them, we deduce that

$$\left(\frac{f_1}{f_2}\right)' = a\frac{f_1}{f_2} + b\frac{1}{f_2} + g, \qquad \left(\frac{1}{f_2}\right)' = l\frac{f_1}{f_2} + m\frac{1}{f_2} + n, \tag{2.7}$$

(a, b, g, l, m, n = const.) and, by substituting in (1.7) we deduce that

$$(a - \varphi l)\frac{f_1}{f_2} + (b - \varphi m)\frac{1}{f_2} + (g - \varphi n) = \frac{\varphi\varphi'' - 2\varphi_2'^2}{\varphi'}\frac{1}{f_2} - \frac{\varphi''}{\varphi'}\frac{f_1}{f_2} + \Phi(v).$$

Since *only* the curve γ is a straight line, this equation must be an identity in f_1/f_2 and $1/f_2$ and consequently

$$\varphi\varphi'' - 2\varphi'^2 = (b - \varphi m)\varphi'; \qquad -\varphi'' = (a - l\varphi)\varphi'$$

from which we deduce successively

$$b - \varphi m = -2\varphi' + \varphi(l\varphi - a),$$
$$2\varphi' = l\varphi^2 + (m - a)\varphi - b,$$
$$2\varphi'' = (2l\varphi + [m - a])\varphi' = 2(l\varphi - a)\varphi'.$$

Since φ is not a constant, $m - a = -2a$, or

$$a + m = 0. \tag{3.7}$$

Our equations become:

$$\varphi' = \tfrac{1}{2}l\varphi^2 + m\varphi - \frac{b}{2},$$
$$\left(\frac{f_1}{f_2}\right)' = -m\frac{f_1}{f_2} + b\frac{1}{f_2} + g, \qquad \left(\frac{1}{f_2}\right)' = l\frac{f_1}{f_2} + m\frac{1}{f_2} + n. \tag{4.7}$$

Let us now consider the collineation

$$\bar{x} = \frac{Ax + By + E}{Lx + My + N}, \qquad \bar{y} = \frac{y}{Lx + My + N} \qquad \begin{pmatrix} A, B, \cdots, N = \text{const.} \\ AN - LE \neq 0 \end{pmatrix}.$$

It transforms the line γ $(y = 0)$ into itself. The point (f_1, f_2) of c is transformed into the point $(\bar{f}_1, \bar{f}_2)$ determined by

$$\frac{\bar{f}_1}{\bar{f}_2} = A\frac{f_1}{f_2} + B + E\frac{1}{f_2}, \qquad \frac{1}{\bar{f}_2} = L\frac{f_1}{f_2} + M + N\frac{1}{f_2}. \tag{5.7}$$

The new function $\bar{\varphi}$ will be given by

$$\bar{\varphi} = \frac{A\varphi + E}{L\varphi + N}, \tag{6.7}$$

and therefore

$$\bar{\varphi}' = \tfrac{1}{2}\left(\bar{l}\bar{\varphi}^2 + \bar{m}\bar{\varphi} - \frac{\bar{b}}{2}\right) \tag{7.7}$$

if

$$\bar{l} = \frac{1}{AN - LE}(lN^2 - 2mLN - bL^2),$$
$$\bar{m} = \frac{1}{AN - LE}(-lEN + m[LE + AN] + bLA),$$
$$\bar{b} = \frac{1}{AN - LE}(-lE^2 + 2mAE + bA^2),$$

from which follows

$$\bar{l}\bar{b} + \bar{m}^2 = lb + m^2.$$

We could prove the same equations by studying (5.7) instead of (6.7). If $lb + m^2 \neq 0$, we can choose the values of $N:L$ and $A:E$ such that $l = \bar{b} = 0$ even if it is necessary (as in our case) that $AN - EL \neq 0$, or $N:L \neq E:A$. If $lb + m^2 = 0$, I can choose $N:L$ such that $l = 0$; since $l\bar{b} + \bar{m}^2 = lb + m^2 = 0$, $\bar{m}$ also will be equal to zero. Therefore, by changing notation and disregarding the dashes, I can always suppose that

$$(8.7) \qquad l = b = 0, \qquad m \neq 0 \quad \text{or} \quad l = m = 0, \qquad b \neq 0.$$

In the latter case I have added the condition $b \neq 0$. If it were not satisfied and l, m, b were equal to zero, from (4.7) we could deduce

$$\left(\frac{f_1}{f_2}\right)' = g, \qquad \left(\frac{1}{f_2}\right)' = n.$$

The line c would be a straight line $f_1/f_2 = \text{const.}$, if $g = 0$, a straight line $f_2 = \text{const.}$, if $n = 0$, or a straight line $g\,1/f_2 - n f_1/f_2 = \text{const.}$, if $ng \neq 0$; which is contrary to our hypothesis that among our three curves only γ is a straight line. I have now to study the two cases (8.7). In the former I can multiply u (and consequently also v) by the same constant factor (which is of no importance for our problem), such that $m = 1$; in the latter I can analogously suppose $b = 1$. Therefore in the first case:

$$\left(\frac{f_1}{f_2}\right)' = -\frac{f_1}{f_2} + g, \qquad \left(\frac{1}{f_2}\right)' = \frac{1}{f_2} + n, \qquad \varphi' = \varphi;$$

in the second case:

$$\left(\frac{f_1}{f_2}\right)' = \frac{1}{f_2} + g, \qquad \left(\frac{1}{f_2}\right)' = n, \qquad \varphi' = -\tfrac{1}{2}.$$

In the first case

$$\frac{f_1}{f_2} - g = Ae^{-u}, \qquad \frac{1}{f_2} + n = Be^{u}, \qquad \varphi = e^{Dv}, \qquad (A, B, D \text{ const.})$$

$$(1 + nf_2)(f_1 - gf_2) = ABf_2^2.$$

The line c is the conic

$$(1 + ny)(x - gy) = ABy^2.$$

By means of (4.2), (2.2), (1.2) we calculate the coordinates $X = F_1$, $Y = F_2$ of a point of C. We find

$$X = F_1 = \frac{-g + BDe^{u+v}}{\frac{1}{D}Ae^{-(u+v)} + n}, \qquad Y = F_2 = \frac{-1}{\frac{1}{D}Ae^{-(u+v)} + n}$$

[both functions of $w = -(u + v)$], and we remark that the point (X, Y) generates the same conic c, so that the curves c, C are identical. There are *two*

distinct points of intersection of the conic $c = C$ and of the line γ ($y = 0$): the point $x = y = 0$ and the point at infinity of γ. The representation of the coordinates of the points of our lines by means of the corresponding parameter requires the *exponential* function. The conic c and the line γ form together a cubic with two double points. In the second case I will apply, for short, such a collineation that $g = 0$. (I have to substitute

$$\bar{f}_1 = \frac{f_1}{gf_2 + 1}, \qquad \bar{f}_2 = \frac{f_2}{gf_2 + 1}, \qquad \bar{\varphi} = \bar{\varphi}_1 = \frac{\varphi_1}{g\varphi_2 + 1} = \varphi_1 = \varphi,$$

$$\bar{\varphi}_2 = \varphi_2 = 0, \qquad \text{for } f_1, f_2, \varphi_1 \varphi_2$$

respectively, and to neglect the dashes afterward.) We find

$$\frac{1}{f_2} = nU, \qquad \frac{f_1}{f_2} = \tfrac{1}{2}nU^2 + B, \qquad \varphi = -\tfrac{1}{2}V,$$

where $U = u + \text{const.}$, $V = v + \text{const.}$, $B = \text{const.}$ Therefore the line c generated by the point $x = f_1$, $y = f_2$ is again a conic

$$2n(xy - By^2) = 1.$$

If $W = -(U + V)$, we find that the point (X, Y) of the third line C is the point

$$X = \frac{W}{2} + \frac{B}{n}\frac{1}{W}, \qquad Y = \frac{1}{n}\frac{1}{W},$$

so that it generates the same conic $c = C$. In this case *rational functions* were sufficient, but the conic c and the line γ ($y = 0$) are now *tangent to each other* (at infinity). In conclusion: *If one of the curves c, γ, C is a straight line, then either the other lines are straight lines too, or they belong to the same conic. The coordinates of their points are rational functions of their parameters if the curves are lines of the same pencil, or are a straight line and a conic tangent to each other; otherwise it is necessary to make use of the exponential function.*

8. General deductions

From now on we can suppose that none of the three lines is a straight line.

I shall now choose as origin a point O of the curve γ. This point has to satisfy the following condition. If γ and c belong to algebraic curves $\bar{\gamma}$ and $\bar{c}$, and these algebraic curves are not identical (or, if c, γ do not belong to the same algebraic curve), the point O must not belong to $\bar{c}$ (is not an intersection of $\bar{c}, \bar{\gamma}$). If the curves c, γ are identical, one obviously cannot satisfy this condition, and O is an arbitrary point of γ. Afterwards I choose as y-axis ($x = 0$) the tangent to γ at O. Therefore $\varphi_1 = \varphi_2 = \varphi_1' = 0$ at O and consequently $\varphi_2' \neq 0$. Since γ is not a straight line I can also suppose that O is not a point of inflection of γ, so that $\varphi_1'' \neq 0$ at O. The reduced equation (2.4) reads

$$\left(\frac{f_2}{f_1}\right)' = A\left(\frac{f_2}{f_1}\right)^2 + l\frac{f_2}{f_1} + m\frac{1}{f_1} + n, \tag{1.8}$$

in which the constants A, l, m, n are the values at O of φ_1''/φ_2', $-\varphi_2''/\varphi_1'$, $-2\varphi_2'$, D_1. Therefore in our case

$$Am \neq 0. \tag{2.8}$$

I write now the reduced equation (8.4) for any other point $\Omega = (\varphi_1, \varphi_2)$ of γ. And I remark that one can deduce from (1.8) that the left-hand member of (8.4) is

$$(f_1 - \varphi_1)f_2' - (f_2 - \varphi_2)f_1' = (f_1 - \varphi_1)\left[f_1\left(\frac{f_2}{f_1}\right)' + \frac{f_2}{f_1}f_1'\right] - (f_2 - \varphi_2)f_1'$$

$$= (f_2 - \varphi_1)(lf_2 + m + nf_1) + (f_2 - \varphi_2)Af_2 + (\varphi_2 f_1 - \varphi_1 f_2)\left(A\frac{f_2}{f_1} + \frac{f_1'}{f_1}\right).$$

The right-hand member of (8.4) is a polynomial of second degree in f_1, f_2. Consequently we deduce from (8.4) that

$$-\frac{1}{f_1^2}(\varphi_2 f_2 - \varphi_1 f_2)\left(A\frac{f_2}{f_1} + \frac{f_1'}{f_1}\right) = -\left(\varphi_2 - \varphi_1\frac{f_2}{f_1}\right)\left[A\frac{1}{f_1}\cdot\frac{f_2}{f_1} - \left(\frac{1}{f_1}\right)'\right]$$

is a polynomial P of second degree in f_2/f_1, $1/f_1$, the coefficients of which are functions of the (arbitrary) value of v corresponding to the (arbitrary) point Ω of γ. Therefore

$$\left(\frac{1}{f_1}\right)' = A\frac{1}{f_1}\frac{f_2}{f_1} + \frac{P\left(\frac{f_2}{f_1}, \frac{1}{f_1}\right)}{\varphi_2 - \varphi_1\frac{f_2}{f_1}}. \tag{3.8}$$

If P were divisible by $\varphi_2 - \varphi_1 f_2/f_1$ the quotient would be a polynomial of first degree:

$$a\frac{f_2}{f_1} + b\frac{1}{f_1} + g. \tag{4.8}$$

Its coefficients a, b, g ought to be independent of v (and therefore constant) because, if it were not so, by equating the values of (4.8) corresponding to two different values of v, we should find an equation of first degree in f_2/f_1, $1/f_1$, and the curve c would be a straight line; which is contrary to our hypotheses. Therefore in this case

$$\left(\frac{f_2}{f_1}\right)' = A\left(\frac{f_2}{f_1}\right)^2 + l\frac{f_2}{f_1} + m\frac{1}{f_1} + n,$$

$$\left(\frac{1}{f_1}\right)' = A\frac{1}{f_1}\frac{f_2}{f_1} + a\frac{f_2}{f_1} + b\frac{1}{f_1} + g, \qquad (A, l, m, n, a, b, g = \text{const.}) \tag{5.8}$$

or

$$-f_1' = f_1(af_2 + gf_1) + bf_1 + Af_2,$$

$$-f_2' = f_2(af_2 + gf_1) - nf_1 + (b - l)f_2 - m. \tag{6.8}$$

If P is not divisible by $\varphi_2 - \varphi_1 f_2/f_1$, we remark that, by giving to v two different values v_1, v_2, we can deduce from (3.8) that

$$\frac{P_1\left(\frac{f_2}{f_1}, \frac{1}{f_1}\right)}{\varphi_2(v_1) - \varphi_1(v_1)\frac{f_2}{f_1}} = \frac{P_2\left(\frac{f_2}{f_1}, \frac{1}{f_1}\right)}{\varphi_2(v_2) - \varphi_1(v_2)\frac{f_2}{f_1}}, \tag{7.8}$$

[P_1 and P_2 are the polynomials deduced from P by supposing $v = v_1$ or $v = v_2$.]

We can deduce from our supposition that (7.8) cannot be an identity in f_2/f_1, $1/f_1$, because the denominators are not proportional to each other. (If they were proportional, the straight lines L_i defined by

$$\varphi_2(v_i) - \varphi_1(v_i)\frac{f_2}{f_1} = 0, \qquad (i = 1, 2)$$

would be the same line; but these lines are the straight lines joining the origin with the points $v = v_1$ and $v = v_2$ of γ; these three points are not collinear, because γ is not a straight line; and therefore L_1, L_2 cannot be identical.) Therefore (7.8) is *an equation* satisfied by all the points of c. By means of a collineation I can suppose

$$x = \frac{1}{f_1}, \qquad y = \frac{f_2}{f_1}. \tag{8.8}$$

The equation (7.8) becomes

$$yS = T \tag{9.8}$$

if

$$S = \varphi_1(v_1)P_2(y, x) - \varphi_1(v_2)P_1(y, x); \qquad T = \varphi_2(v_1)P_2 - \varphi_2(v_1)P_1.$$

The S, T are polynomials in x, y of a degree not greater than two. The curve c will satisfy the equation (8.8) and the equations [deduced from (1.8) and (3.8)]

$$\begin{aligned} y' &= Ay^2 + ly + mx + n, \\ x' &= Axy + \frac{Q(y, x)}{y - \eta}. \end{aligned} \tag{10.8}$$

[Q is a polynomial in x, y of a degree not greater than 2. Its coefficients are functions of v and $\eta = \varphi_2/\varphi_1$.] We have consequently to study two systems of equations: the system (5.8) and the system (9.8) and (10.8).

9. The equations (5.8) or (6.8)

These equations demonstrate that the first member of (8.4) is equal to

$$\begin{aligned} &[\varphi_1(f_2 - \varphi_2) - \varphi_2(f_1 - \varphi_1)](af_2 + gf_1) + (f_2 - \varphi_2)(bf_1 + Af_2) \\ &\qquad + (f_1 - \varphi_1)\{nf_1 + (l - b)f_2 + n\}. \end{aligned} \tag{1.9}$$

This quantity must be equal to the second member of (8.4). We obtain in this way an equation in f_1, f_2 (of a degree not greater than two). I demonstrate that I can suppose that this is an *identity* in f_1, f_2. If one at least of the curves c, γ, C does not belong to a conic, I can choose this curve as curve c. Since c is neither a conic nor a straight line (according to our actual hypothesis), its points cannot satisfy an equation of a degree not greater than two; and therefore the preceding equation is an identity in f_1, f_2. If every one of the arcs c, γ, C belongs to a conic, these three conics cannot be identical to each other, because it is not possible that three points of the same conic are collinear. Therefore at least two of these conics are different from each other; and I can suppose that c, γ belong to different conics. But, by writing that (1.9) is equal to the second member of (8.4) I obtain an equation of second degree in $x = f_1$, $y = f_2$, which is satisfied when we suppose that $x = \varphi_1$ $y = \varphi_2$.[5] If this equation were not an identity, it would be the equation of a conic, to which the point $x = \varphi_1$, $y = \varphi_2$ belongs; but this point is any point whatsoever of γ; and therefore every point of γ would lie on the same conic to which c belongs; which is contrary to our supposition. Therefore (1.9) and the second member of (8.4) are identically equal to each other. I order both according to powers of $f_1 - \varphi_1$, $f_2 - \varphi_2$; by comparing the coefficients of similar terms, I conclude that

$$D\varphi_1'^2 = a\varphi_1 + A, \qquad \frac{\varphi_1''}{\varphi_1'} - 2D\varphi_1'\varphi_2' = l + g\varphi_1 - a\varphi_2,$$

$$-\frac{\varphi_1''\varphi_2' + \varphi_2''\varphi_1'}{2(\varphi_1')^2} + D\varphi_2'^2 = -g\varphi_2 + n, \tag{2.9}$$

$$-2\varphi_2' = -\varphi_2(a\varphi_2 + g\varphi_1) + m + n\varphi_1 + (l - b)\varphi_2,$$

$$2\varphi_1' = \varphi_1(a\varphi_2 + g\varphi_1) + b\varphi_1 + A\varphi_2,$$

and, by eliminating D from the first three equations:

$$\varphi_1'' = 2\varphi_2'(A + a\varphi_1) + (l + g\varphi_1 - a\varphi_2)\varphi_1', \tag{3.9}$$

$$\varphi_2'' = \varphi_2'(a\varphi_2 - g\varphi_1 - l) + 2(g\varphi_2 - n)\varphi_1'.$$

By comparing (3.9) with the values of φ'' deduced from (2.9) by differentiating, we obtain

$$(b - 2l + 3a\varphi_2)\varphi_1' - 3(A + a\varphi_1)\varphi_2' = 0, \tag{4.9}$$

$$(b + l + 3g\varphi_1)\varphi_2' + 3(n - g\varphi_2)\varphi_1' = 0. \tag{5.9}$$

If $a \neq 0$, the equation (4.9) gives

$$\frac{\varphi_2 + \dfrac{b - 2l}{3a}}{\varphi_1 + \dfrac{A}{a}} = \text{const.},$$

[5] This would be contrary to what we supposed at the beginning of §8.

and the line γ would be a straight line, which is now absurd. Therefore $a = 0$, and the equation (4.9) proves that

$$(b - 2l)\varphi_1 - 3A\varphi_2 = \text{const.}$$

Since γ is not a straight line, we deduce

$$b - 2l = 3A = 0.$$

(We do not now take into account that $A \neq 0$ [according to (2.8)].) If $g \neq 0$, we deduce from (5.9)

$$\frac{\varphi_1 + \dfrac{b + l}{3g}}{\varphi_2 - \dfrac{n}{g}} = \text{const.},$$

and we deduce, as above, that $g = 0$; and, as above, (5.9) proves that $b + l = n = 0$. Therefore $A = a = g = b = l = n = 0$; and therefore, from (2.9) we deduce $\varphi_1 = \text{const.}$, which is absurd because γ is not a straight line. Consequently we can disregard the equations (5.8) and (6.8).

10. The equations (9.8) and (10.8)

I have already supposed that none of the three lines c, γ, C is a straight line. If every one of the lines c, γ, C belongs to a conic, I have already remarked that these three conics cannot be identical, and that I am therefore allowed to suppose that the conics to which c and γ belong are not identical. In this case I am therefore allowed to choose as origin $O(\varphi_1 = \varphi_2 = 0)$ a point of γ which does not lie on the conic to which c belongs. If

$$A_{11}f_1^2 + 2A_{12}f_1f_2 + A_{22}f_2^2 + 2A_{13}f_1 + 2A_{23}f_2 + A_{33} = 0 \quad (A_{ij} = \text{const.})$$

is the equation of the conic to which c belongs, we can therefore admit that $A_{33} \neq 0$. By means of the collineation (8.8) the equation of this conic is transformed into the equation

$$T = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0 \tag{1.10}$$

$$(a_{11} = A_{33}, \quad a_{12} = A_{23}, \text{ etc.}),$$

in which

$$a_{11} = A_{33} \neq 0. \tag{2.10}$$

The equations (10.8) can be written in the form

$$y' = Ay^2 + ly + mx + n, \qquad x' = Axy + \frac{Q(y, x)}{y - \eta}, \tag{3.10}$$

in which $\eta = \varphi_2/\varphi_1$, and Q is a polynomial in y, x (of a degree not greater than two), the coefficients of which are functions of v only.

By differentiating (1.10) and taking into account (3.10) we find

$$(a_{11}x + a_{12}y + a_{13})\left(Axy + \frac{Q(y, x)}{y - \eta}\right) + (a_{21}x + a_{22}y + a_{23})(Ay^2 + ly + mx + n) = 0,$$

or

$$(y - \eta)\left[Axy + \frac{a_{21}x + a_{22}y + a_{23}}{a_{11}x + a_{12}y + a_{13}}(Ay^2 + ly + mx + n)\right] = -Q(x, y). \tag{4.10}$$

By giving to v two different values, we obtain two equations; if η_i, Q_i $(i = 1, 2)$ are the corresponding values of η and of Q, we find by subtracting that

$$Axy + \frac{a_{21}x + a_{22}y + a_{23}}{a_{11}x + a_{12}y + a_{13}}(Ay^2 + ly + mx + n) = \frac{Q_1 - Q_2}{\eta_1 - \eta_2} = R,$$

where R is a polynomial in x, y, of a degree not greater than two, with constant coefficients. From (4.10) we can deduce that yR also is, on the curve c belonging to the conic $T = 0$, equal to such a polynomial, and therefore we can find an identity (in x, y)

$$yR = (px + qy + r)T + \text{(a polynomial of a degree not greater than two)}$$

$(p, q, r = \text{const.})$. By comparing the terms in x^3, and by recalling that $a_{11} \neq 0$, we find $p = 0$; and, since the preceding equation is an identity, we deduce that

$$R = qT + \text{(a polynomial of first degree in } x, y).$$

Therefore, on the curve c (which satisfies the equation $T = 0$) R is equal to a polynomial of first degree in x, y; its coefficients must be constant, because γ is not a straight line. And therefore we deduce that, on c,

$$\frac{Q}{y - \eta} = -R$$

can be considered equal to a polynomial of first degree in x, y with constant coefficients. The equations (3.10) are therefore identical with equations which we have already studied; and therefore we can suppose, from now on, that *none* of the three curves c, γ, C belongs to a straight line, and that *at least one* of them does not belong to a conic.

11. The equations (9.8) and (10.8) in the general case

We can also suppose that c does not belong to a conic, and consequently the equation (9.8) is precisely of third degree. If

$$yS_1 = T_1 \qquad (S_1, T_1 \text{ polynomials in } x, y)$$

is another equation of third degree, satisfied by the points of c, this equation must be identical to (9.8) up to a constant factor k so that

$$y(S_1 - kS) = T_1 - kT \qquad \text{(identically)}.$$

Therefore $T_1 - kT$ must be divisible by y; if $Lx + My + N$ is the quotient (L, M, N = const.), we conclude that

$$S_1 = kS + Lx + My + N, \qquad T_1 = kT + y(Lx + My + N).$$

Since x' is independent of v, we deduce from (10.8) that $Q/(y - \eta)$ is independent of v. We give to v two new values v_3, v_4, and suppose that Q_3, Q_4 are the polynomials deduced from Q in this manner. We deduce that

$$Q_3[y - \eta(v_4)] = Q_4[y - \eta(v_3)]$$

is another non-identical equation satisfied by the points of c. We deduce therefore that

$$Q_3 - Q_4 = kS + Lx + My + N,$$

$$Q_3\eta(v_4) - Q_4\eta(v_3) = kT + y(Lx + My + N),$$

in which k, L, M, N depend on v_3, v_4. Therefore, by recalling that $yS = T$ on the curve c, we find that, on the curve c:

$$Q_3 = \frac{1}{\eta(v_4) - \eta(v_3)}\{k[T - S\eta(v_3)] + [y - \eta(v_3)](Lx + My + N)\},$$

$$\frac{Q_3}{y - \eta(v_3)} = \frac{1}{\eta(v_4) - \eta(v_3)}\{kS + Lx + My + N\},$$

which must be independent of both v_3, v_4 [see (10.8)]. Therefore, since c does not belong to a conic,

$$\frac{k}{\eta(v_4) - \eta(v_3)}, \qquad \frac{L}{\eta(v_4) - \eta(v_3)}, \qquad \frac{M}{\eta(v_4) - \eta(v_3)}, \qquad \frac{N}{\eta(v_4) - \eta(v_3)}$$

must be constant, and therefore

$$\frac{Q}{y - \eta} = \frac{Q_3}{y - \eta(v_3)} = hS + px + qy + r \qquad (h, p, q, r = \text{const.}).$$

I shall write

$$\overline{S} = hS + px + qy + r, \qquad \overline{T} = hT + y(px + qy + r).$$

The equations (9.8), (10.8) become:

$$y\overline{S} = \overline{T}; \qquad y' = Ay^2 + ly + mx + n, \qquad x' = Axy + \overline{S}$$

or, by changing notation and disregarding the dashes:

(1.11) $$yS = T,$$

(2.11) $$y' = Ay^2 + ly + mx + n, \qquad x' = Axy + S.$$

According to our hypotheses (§8) the origin ($\varphi_1 = \varphi_2 = 0$) does not belong to the cubic $yS = T$, if the curve γ does not belong to this cubic. We have applied a projective transformation (8.8); if we make use of homogeneous projective coordinates x_i such that $x_1:x_2:x_3 = x:y:1$, this transformation is defined by

$$x_1:x_2:x_3 = 1:f_2:f_1 \qquad \text{(for the points of } c\text{)},$$

and the analogous equations

$$\xi_1:\xi_2:\xi_3 = 1:\varphi_2:\varphi_1 \qquad \text{(for the points of } \gamma\text{)}.$$

The equation (1.11) is turned into an equation

$$x_2 \sum a_{ik}x_ix_k = x_3 \sum t_{ik}x_ix_k\,, \tag{3.11}$$

(a_{ik} and t_{ik} are the coefficients of S, T) which is satisfied by the point $x_2 = x_3 = 0$ (which is the transform of the former origin). Therefore the initial origin lies on our cubic; and consequently all the curve γ belongs, like c, to this cubic.

It may perhaps be interesting to prove directly that the curves c, γ, have the same tangent at the origin $O(x_2 = x_3 = 0)$. The tangent to γ at this point was the line $\varphi_1 = 0$, which is now defined by $x_3 = 0$. By writing $x_3 = 0$ in the equation (3.11) of c, I find

$$x_2(a_{11}x^2 + 2a_{12}xy + a_{22}y^2) = 0.$$

We shall have proved that the line $x_3 = 0$ is tangent also to c, if we prove that $a_{11} = 0$. By differentiating the first of the equations (1.11) we find [by taking into account the values of x', y' given by (2.11)] that, on the curve c,

$$\begin{aligned} 0 = {} & [2a_{11}xy + 2a_{12}y^2 + 2a_{13}y - 2(t_{11}x + t_{12}y + t_{13})](Axy + S) \\ & + [a_{11}x^2 + 4a_{12}xy + 3a_{22}y^2 + 2a_{13}x + 4a_{23}y + a_{33} - 2(t_{21}x + t_{22}y + t_{23})] \\ & [Ay^2 + ly + mx + n]. \end{aligned}$$

Therefore the polynomial of fourth degree, (on the right member), must be equal to the product of $yS - T$ by a polynomial p of first degree in x, y. By comparing the terms of fourth degree in the preceding polynomials and the terms of third degree in $yS - T$, we conclude easily that

$$p = 2a_{11}x + (3A + 2a_{12})y + s \qquad (s = \text{const.}).$$

By comparing the coefficients of x^3, we deduce

$$-2t_{11}a_{11} + a_{11}m = -2a_{11}t_{11}\,,$$

and consequently $a_{11} = 0$, because $m \neq 0$ [see (2.8)]. The right-hand members of (2.11) are polynomials in x, y of a degree not greater than two.

What happens if we make a change of projective not-homogeneous coordinates? If

$$\bar{x} = \frac{Bx + Dy + F}{Px + Qy + R}, \qquad \bar{y} = \frac{Lx + My + N}{Px + Qy + R} \tag{4.11}$$

(B, D, $\cdots$, Q, R constants, the determinant of which is different from zero) are the new coordinates, we find, for instance:

$$(5.11) \qquad \bar{x}' = \frac{(BQ - DP)(x'y - xy') + (RB - FP)x' + (RD - FQ)y'}{(Px + Qy + R)^2}.$$

From (2.11) we deduce that x', y' and

$$x'y + y'x = yS - x(lx + my + n) = T - x(lx + my + n),$$

and therefore also the numerators of the right members of (5.11) are, on our curve c, equal to polynomials in x, y, the degree of which is not greater than 2. But the equations (4.11) can be written in the form

$$\bar{x} = B\frac{x}{Px + Qy + R} + D\frac{y}{Px + Qy + R} + F\frac{1}{Px + Qy + R},$$

$$\bar{y} = L\frac{x}{Px + Qy + R} + M\frac{y}{Px + Qy + R} + N\frac{1}{Px + Qy + R},$$

$$1 = P\frac{x}{Px + Qy + R} + Q\frac{y}{Px + Qy + R} + R\frac{1}{Px + Qy + R},$$

and we can deduce from these equations that

$$\frac{x}{Px + Qy + R}, \qquad \frac{y}{Px + Qy + R}, \qquad \frac{1}{Px + Qy + R}$$

are linear integral functions of $\bar{x}$, $\bar{y}$. Therefore the right-hand member of (5.11) is a polynomial in $\bar{x}$, $\bar{y}$ of a degree not higher than two; and we can obtain an analogous result for $\bar{y}'$. Therefore: *For any system of not-homogeneous projective coordinates x, y, their derivatives x', y' are (on c) equal to polynomials in x, y of a degree not higher than two.*

12. The canonical equation of the cubic

I can now choose such not-homogeneous projective coordinates that the equation of the cubic is

$$(1.12) \qquad y^2 = 4x^3 + 2gx + r \qquad (g, r = \text{const.})$$

(according to Weierstrass, $2g = -g_2$, $r = -g_3$). If

$$(2.12) \qquad 8g^3 + 27r^2 \neq 0,$$

the cubic is not rational and has no double point. According to what we have already proved, we can write

$$(3.12) \qquad x' = P_2 + P_1 + P_0, \qquad y' = Q_2 + Q_1 + Q_0,$$

(P_s and Q_s are *homogeneous* polynomials of degree s in x, y) ($s = 0, 1, 2$). By differentiating (1.12) we deduce that

$$(4.12) \qquad (6x^2 + g)(P_2 + P_1 + P_0) - y(Q_2 + Q_1 + Q_0)$$

must be equal to zero at the points of the cubic, and that therefore the polynomial (4.12) must be identically equal to the product

$$(ax + by + k)(4x^3 + 2gx + r - y^2) \tag{5.12}$$

if we choose the constants a, b, k in a suitable manner. If I compare in (4.12) and (5.12) the terms of fourth and third degree, I obtain:

$$P_2 = \tfrac{2}{3}x(ax + by), \qquad 6x^2P_1 - yQ_2 = 4kx^3 - y^2(ax + by). \tag{6.12}$$

From the latter equation one deduces:

$$P_1 = \tfrac{2}{3}kx + py; \quad Q_2 = 6px^2 + y(ax + by) \quad (p = \text{an unknown constant}). \tag{7.12}$$

By comparing the terms of second degree in (4.12) and (5.12), one finds:

$$6x^2P_0 - yQ_1 + gP_2 = 2gx(ax + by) - cy^2.$$

The value of P_2 is given by (6.12); we deduce

$$P_0 = \tfrac{2}{9}ga, \qquad Q_1 = ky - \tfrac{4}{3}gbx. \tag{8.12}$$

By comparing the other terms of (4.12) and (5.12), we obtain

$$gP_1 - yQ_0 = r(ax + by) + 2kgx, \qquad gP_0 = kr,$$

and from (7.12), (8.12) we deduce:

$$Q_0 = gp - rb, \tag{9.12}$$

$$ra + \tfrac{4}{3}kg = 0, \qquad 2g^2a = 9kr. \tag{10.12}$$

These last equations are linear in a, k; if (2.12) is satisfied and the cubic is not rational, we deduce $a = k = 0$, and therefore

$$x' = P_2 + P_1 + P_0 = y(\tfrac{2}{3}bx + p), \tag{11.12}$$

$$\begin{aligned} y' = Q_2 + Q_1 + Q_0 &= 6px^2 + by^2 - \tfrac{4}{3}gbx + gp - rb \\ &= (6x^2 + g)(\tfrac{2}{3}bx + p). \end{aligned} \tag{12.12}$$

Therefore, if

$$U = \int \frac{dx}{y} \tag{13.12}$$

is the Abelian integral of first kind connected with our cubic, we deduce that

$$\frac{dU}{du} = \frac{2}{3}bx + p. \tag{14.12}$$

But the equation (8.4) is invariant under the group of collineations; I can consequently write it by substituting x, y, ξ, η for f_1, f_2, φ_1, φ_2 respectively. It proves that, on c

$$(x - \xi)y' - (y - \eta)y'$$

is equal to a polynomial of second degree in x, y. Since y' and x', and consequently also $\xi y'$, $\eta x'$ are equal to such polynomials, $xy' - yx'$ also is a polynomial of second degree in x, y. From (11.12) and (12.12) we deduce that on our cubic

$$(\tfrac{2}{3}bx + p)(6x^3 + gx - y^2) = (\tfrac{2}{3}bx + p)(\tfrac{1}{2}y^2 - gx - r)$$

is such a polynomial; and this can be true only if $b = 0$. The equation (14.12) proves that u can also be considered as the Abelian integral of first kind (for our cubic), because this integral is defined up to a multiplicative and an additive constant.

We can now develop the same consideration for the curve γ, which is identical with c; and we should find that the corresponding parameter v is given by

$$\frac{dV}{dv} = q \qquad (q = \text{const.}).$$

[See the analogous equations (13.12) and (14.12).] If $p = q$ our theorem is completely demonstrated, because it is a consequence of Abel's theorem; the third curve is identical with c, γ; the corresponding parameters are the same integral of first kind. Consequently we have to prove only that $p = q$. If $p = q$, our theorem is true and therefore the reduced equation (9.4) is satisfied, even, if we substitute, as before, x, y, ξ, η for f_1, f_2, φ_1, φ_2. If our problem could be solved also for another value of $p \neq q$, the equation (9.4) would be satisfied also if we write $(q/p)\,du$ for du, and if we change, if necessary, the function D. By subtracting the two equations which we deduce in this way from (9.4) we obtain that

$$\frac{d}{du}\frac{x-\xi}{\Omega} = \frac{d}{du}\frac{x-\xi}{\eta'(x-\xi) - \xi'(y-\eta)}$$

is only a function of v. Since $(x - \xi)/\Omega$ cannot be independent of u, because c is not a straight line, we would find, by integrating, that u is a linear (non-integer) function of x, y; which is absurd. Our theorem is therefore completely proved.

13. The cubic is rational and possesses a double point

Let us now suppose that (2.12) is not satisfied and that the cubic possesses a double point $x = \rho$, $y = 0$. The equation of the cubic is now

$$y^2 = 4(x-\rho)^2(x+2\rho) \qquad (g = -2\rho^2;\ r = 8\rho^3). \tag{1.13}$$

At first we suppose $\rho \neq 0$. The equations (10.12) state only that $k = a\rho$ so that we obtain

$$x' = \tfrac{2}{3}(ax + by) + \tfrac{2}{3}a\rho x + py - \tfrac{4}{3}a\rho^2,$$

$$y' = 6px^2 + y(ax + by) + a\rho y + 8\rho^2 bx - 6\rho^2 p - 8\rho^3 b.$$

By writing that $xy' - yx'$ is, on the cubic, equal to a polynomial of second degree in x, y, we find that

$$6px^3 + \tfrac{1}{3}xy(ax + by)$$

must be equal to such a polynomial if we take into account the equation of c. Therefore $a = b = 0$,

$$x' = py, \qquad y' = 6p(x^2 - \rho^2), \qquad du = \frac{1}{p}\frac{dx}{y}, \tag{2.13}$$

which is completely analogous to the definitions (13.12) of the integral of first kind for not rational cubics. In the same manner we shall find, for the parameter v corresponding to the curve γ

$$dv = \frac{1}{q}\frac{d\xi}{\eta} \qquad (q = \text{const.}). \tag{3.13}$$

As above, it will be sufficient to demonstrate that our problem is solved by supposing $p = q$; which is no longer a consequence of Abel's theorem. By introducing a new parameter

$$m = \frac{y}{2(x - \rho)}$$

we find that

$$x = m^2 - 2\rho, \qquad y = 2m(m^2 - \rho), \tag{4.13}$$

are new parametric equations of our cubic. We also prove easily that three points of the cubic, corresponding to the values m, μ, M of the new parameter, are collinear if and only if

$$Mm + M\mu + m\mu = -3\rho. \tag{5.13}$$

From (2.13), (4.13) it follows that

$$u = \frac{1}{p}\int \frac{dm}{m^2 - 3\rho}, \quad \text{or} \quad m = -\sqrt{3\rho}\,\frac{e^v + 1}{e^v - 1} \tag{6.13}$$

$$[U = 2p\sqrt{3\rho}\,u + h].$$

[The constant h is arbitrary.] In the same manner we find that the value μ of the new parameter can be obtained from the corresponding value of v by means of the equation

$$\mu = -\sqrt{3\rho}\,\frac{e^V + 1}{e^V - 1}, \qquad [V = 2p\sqrt{3\rho}\,v + l], \qquad (l = \text{const.}). \tag{7.13}$$

The equations (5.13), (6.13), (7.13) give

$$M = -\sqrt{3\rho}\,\frac{e^W + 1}{e^W - 1}$$

if $W = -(U + V)$. We have found the condition for the parameters, which was our starting point. Our theorem is proved also in this case; the three curves c, γ, C are identical. The introduction of the parameters U, V, W or u, v, w, requires, in this case, only the exponential function.

14. The cubic is rational and possesses a cusp

In this case $\rho = 0$; the equation of the cubic is

$$y^2 = 4x^3 \qquad (g,\ r = 0). \tag{1.14}$$

The equations (10.12) in a, k are identities. From the results of §12 and from the remark that $xy' - yx'$ must be, on the cubic c, equal to a polynomial of second degree in x, y, we deduce that

$$x' = \tfrac{2}{3}kx + py, \qquad y' = 6px^2 + ky. \tag{2.14}$$

We can write the reduced equation (7.4) by substituting x, y, ξ, η for f_1, f_2, φ_1, φ_2, respectively. From (2.14) we conclude (by taking into account that the equation of γ is $\eta^2 = 4\xi^3$ because γ is identical with c that the first member of (7.4) is

$$\begin{aligned} z_1 z_2' - z_2 z_1' &= \frac{x-\xi}{\xi'}\frac{y'}{\eta'} - \frac{y-\eta}{\eta'}\frac{x'}{\xi'} \\ &= \frac{1}{\xi'\eta'}\{(x-\xi)(6px^2+ky) - (y-\eta)(\tfrac{2}{3}kx+py)\} \\ &= \frac{p}{\xi'\eta'}\{\tfrac{1}{2}(\eta' z_2)^2 + 2\eta\eta' z_2 - 6\xi\xi'^2 z_1^2 - 12\xi^2\xi' z_1\} \\ &\quad + k\frac{1}{\xi'\eta'}\{\tfrac{1}{3}\xi'\eta' z_1 z_2 + \eta\xi' z_1 - \tfrac{2}{3}\xi\eta' z_2\} \\ &\qquad \left(z_1 = \frac{x-\xi}{\xi'},\ z_2 = \frac{y-\eta}{\eta'}\right). \end{aligned}$$

By comparing the coefficients of $2z_1$, $2z_2$ in the two members of (7.4), I obtain

$$1 = p\frac{\eta}{\xi'} - \tfrac{1}{3}k\frac{\xi}{\xi'},$$

$$-1 = -6p\frac{\xi^2}{\eta'} + \tfrac{1}{2}k\frac{\eta}{\eta'}.$$

But $\eta/\xi' = 6\xi^2/\eta'$ is a consequence of the equation of γ; therefore, by summing we obtain $k = 0$, and consequently

$$x' = py, \qquad y' = 6px^2.$$

We find again

$$du = \frac{1}{p}\frac{dx}{y},$$

which is completely analogous to the definition of u in the other cases.

I introduce a new parameter $m = y/2x$ and find that

$$x = m^2, \qquad y = 2m^3$$

are the new parametric equations of c, and that

$$u = -\frac{1}{pm} + h \qquad (h = \text{const.}).$$

The introduction of u requires in this case only *rational* functions. As above, we have only to demonstrate that, by supposing

$$v = -\frac{1}{p\mu} + l \qquad (l = \text{const.}),$$

our conditions are satisfied. Now the three points of c, corresponding to the values m, μ, M of the new parameter, are collinear if and only if

$$\frac{1}{m} + \frac{1}{\mu} + \frac{1}{M} = 0,$$

or if

$$u + v + w = 0,$$

in which we have written

$$w = -\frac{1}{pM} - (l + h).$$

Consequently our theorem is demonstrated in every case.

Conclusions

It is possible to state a theorem which may be considered as Abel's converse theorem for a cubic, but, besides the general case which leads to elliptic functions, we have to consider many particular cases:

1. The cubic degenerates into three straight lines forming a triangle,
2. The cubic degenerates into three straight lines belonging to the same pencil,
3. The cubic degenerates into a conic and a straight line, which intersect at two distinct points,
4. The cubic degenerates into a conic and a straight line, which are tangent to each other,
5. The cubic is rational and possesses a double point,
6. The cubic is rational and possesses a cusp.

In cases 1, 3, 5, the introduction of the corresponding parameters requires the *exponential* function, whereas *rational* functions are sufficient in cases 2, 4, 6.

Generalizations

It is obvious that the preceding results can be easily generalized to curves or a degree higher than 1. For instance we can suppose we have six curves (or

small arcs of curves) c_1 , c_2 , $\cdots$, c_6 ; the coordinates of a point of c_i are functions of a parameter u_i . I choose on every curve c_i a point A_i , and I suppose that these points lie on a conic, if and only if the sum of the corresponding parameters u_i is equal to zero. To study this problem I have to consider the conics which degenerate into two straight lines, and to choose three curves arbitrarily among the six curves c_i ,—for instance c_1 , c_2 , c_3 . I choose three *collinear* points A_1 on c_1 , A_2 on c_2 , A_3 on c_3 , and consider the points of c_4 , c_5 , c_6 and the corresponding parameters u_4 , u_5 , u_6 as variable. Our problem becomes: When does it happen that three points A_4 of c_4 , A_5 of c_5 , A_6 of c_6 are collinear, if

$$u_4 + u_5 + u_6 = -(u_1 + u_2 + u_3) = \text{const.?}$$

From our preceding results we deduce that c_4 , c_5 , c_6 form together a cubic; Therefore *three curves, chosen arbitrarily among the six given curves, form a cubic.* and it is very easy to discuss completely this new problem. We remark also that the problem is no more general if we suppose $\sum \Phi_i(u_i) = 0$ (Φ_i a function of u_i only). It may be reduced to the preceding problem by means of a change of parameters.

Another Generalization

Let us suppose we have four arcs of curves c_1 , c_2 , c_3 , c_4 . We shall consider four functions u_1 , u_2 , u_3 , u_4 of the points A_1 , A_2 , A_3 , A_4 of the curves $c_1 \cdots c_4$ respectively. It is possible to choose these parameters u_i in such a way that, when the points A_i are collinear, then

$$u_1 + u_2 + u_3 + u_4 = 0?$$

Let us suppose that the choice of these parameters is possible. We can state the question: Is this choice completely determined (up to non-essential constants)? If not, in how many ways can we choose linearly independent systems of such parameters u? When we have *three* (linearly independent) systems of parameters u, our question is identical with Lie's problem on the surfaces of translation. But if one disregards for a moment these surfaces, it seems to me that the above-stated general problems and these generalizations also are very interesting from the point of view of one who studies Abel's theorem.

Institute for Advanced Study

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

HAUSDORFF INTEGRAL TRANSFORMATIONS

BY H. L. GARABEDIAN

(Received February 6, 1942)

1. Introduction

It is the object of this paper to study the integral transformation

$$v(x) = \int_0^x u(y)\, d\phi(y/x), \tag{1.1}$$

where $u(x)$ is bounded and continuous, $x \geqq 0$, where $\phi(x)$ (called the *mass function* of the transformation) is a Hausdorff mass function (*vide infra*), and where the integration is in the sense of Riemann-Stieltjes. The transformation is said to be *regular* if the existence of $\lim_{x\to\infty} u(x)$ implies the existence of $\lim_{x\to\infty} v(x)$ and the equality of the limits.

A mass function $\phi(x)$ is said to be a Hausdorff mass function when

(1.2)
(i) $\phi(x)$ is of bounded variation on the interval $0 \leqq x < 1$,
(ii) $\phi(x)$ is continuous at $x = 0$, and $\phi(0) = 0$,
(iii) $\phi(1) = 1$,
(iv) $\phi(x) = \frac{1}{2}[\phi(x - 0) + \phi(x + 0)]$ if $0 \leqq x \leqq 1$.

In what follows we shall use the symbols $(H, \phi(x))$ and $[H, \phi(x)]$ to designate integral and matrix transformations respectively involving the mass function $\phi(x)$.[1]

The transformation (1.1) may be written in the form

$$v(x) = [1 - \phi(1 - 0)]\, u(x) + \int_0^x u(y)\, d\phi(y/x), \tag{1.3}$$

where a possible discontinuity of $\phi(x)$ at $x = 1$ has been removed from the integral in (1.1). In 1924 Silverman [1] studied the transformation (1.3) with the following restrictions on $\phi(x)$:

(1.4)
(i) $\phi(x)$ is continuous, $0 \leqq x \leqq 1$,
(ii) $\phi'(x)$ is continuous, $0 < x \leqq 1$,
(iii) $\phi(0) = 0$,
(iv) $\int_0^1 |\phi'(x)|\, dx \leqq M$.

[1] A discussion of the relationship of mass functions to matrix transformations may be found in [5].

In connection with these conditions it is understood that $\phi(1 - 0) = \phi(1)$. Silverman termed a function $\phi(x)$ satisfying the conditions (1.4) *absolutely regular*. Silverman proved that the transformation (1.3) is regular when $\phi(x)$ is absolutely regular,[2] and also established a condition in the form of an integral equation in order that $(H, \phi_a(x)) \supset (H, \phi_b(x))$, where $\phi_a(x)$ and $\phi_b(x)$ are absolutely regular mass functions.

In a recent paper [2] this writer clarified and interpreted the results of Silverman in the light of certain recent developments ([3] and [4]) in the field of Hausdorff matrix transformations, and thus made significant extensions of Silverman's results on inclusion and equivalence relations among Hausdorff integral transformations.

In this paper we extend Silverman's results to include a much wider class of mass functions than the absolutely regular mass functions just defined.

2. Regularity of $(H, \phi(x))$ when $\phi(x)$ is a Hausdorff mass function

This section is devoted primarily to a proof of the theorem which follows.

THEOREM 1. *A necessary and sufficient condition that the integral transformation $(H, \phi(x))$ be a regular transformation is that $\phi(x)$ satisfy the conditions* (1.2).

Let us first prove that if $\phi(x)$ satisfies the conditions (1.2) then the transformation (1.1) is a regular transformation. Assuming that $\lim_{x\to\infty} u(x) = l$, we wish to prove that $\lim_{x\to\infty} v(x) = l$.

We observe that

$$\int_0^x d\phi(y/x) = \int_0^1 d\phi(s) = \phi(1) - \phi(0) = 1.$$

Then we may write

$$v(x) - l = \int_0^x [u(y) - l]\, d\phi(y/x).$$

Now, choose p so large that $| u(x) - l | < \epsilon$, $x \geqq p$. Hold p fixed and denote by M a number greater than $| u(x) - l |$ in $0 \leqq x \leqq p$. Then, for $x > p$:

$$| v(x) - l | \leqq \int_0^p | u(y) - l | \, | d\phi(y/x) | + \int_p^x | u(y) - l | \, | d\phi(y/x) |$$

$$\leqq M \int_0^p | d\phi(y/x) | + \epsilon \int_p^x | d\phi(y/x) |.$$

We observe that

$$\int_p^x | d\phi(y/x) | \leqq \int_0^x | d\phi(y/x) | = \int_0^1 | d\phi(s) | = V,$$

[2] We note also that Silverman required only that $u(x)$ be bounded and integrable, $0 \leqq x \leqq x_1$. In this paper we have to require the boundedness and continuity of $u(x)$, $x \geqq 0$, owing to difficulties arising from the non-existence of the Stieltjes integral $\int_0^1 f(x)\, dg(x)$, in the case of common discontinuities of $f(x)$ and $g(x)$.

where V is the total variation of $\phi(x)$ in the interval (0, 1). Moreover, since $\phi(x)$ is continuous at $x = 0$ we can find a number X so large that

$$\int_0^p |\, d\phi(y/x)\,| = \int_0^{p/x} |\, d\phi(s)\,| < \epsilon, \qquad x > X.$$

Now, let X' be the greater of the numbers p and X. Then, we have

$$|\, v(x) - l\,| < \eta, \qquad x > X',$$

where $\eta = (M + V)\epsilon$, whence

$$\lim_{x\to\infty} v(x) = l.$$

Finally, we show that the conditions (1.2) are necessary for the regularity of the transformation (1.1). Here we assume that $\lim_{x\to\infty} v(x) = l$ and obtain the conditions (1.2).

We observe first of all that the existence of the integral in (1.1) implies that $\phi(x)$ be a function of bounded variation on the interval $0 \leqq x \leqq 1$.

If in (1.1) we set $u(y) \equiv 1$, we get

$$v(x) = \int_0^x d\phi(y/x) = \phi(1) - \phi(0).$$

For regularity we must have $\phi(1) - \phi(0) = 1$. If we take $\phi(0) = 0$, then $\phi(1) = 1$.

Suppose now that $\phi(x)$ has a discontinuity at $x = 0$, that is, $\phi(+0) = \Delta$, $\phi(0) = 0$. Then, let us define the function

$$\psi(x) = \begin{cases} 0, & x = 0, \\ \Delta, & x > 0. \end{cases}$$

If we put $\phi^* = \phi - \psi$, then $\phi^*(0) = 0$ and $\phi(x)$ is continuous at $x = 0$. Note that $\phi^*(1) = 1 - \Delta$. Now we write

$$v(x) = (1 - \Delta) \int_0^x u(y)\, d\theta(y/x) + \int_0^x u(y)\, d\psi(y/x), \tag{2.1}$$

where we set $\theta(x) = \phi^*(x)/(1 - \Delta)$, so that $\theta(1) = 1$. We suppose now that $u(x)$ is any continuous function, $x \geqq 0$, so that $\lim_{x\to\infty} u(x) = l$. Since $\theta(x)$ satisfies the conditions (1.2), the first integral in (2.1) defines a regular transformation and we have

$$\lim_{x\to\infty} v(x) = l(1 - \Delta) + \Delta \cdot u(0).$$

Now, we choose $u(0) = 0$. Hence, a necessary condition for regularity is the requirement $\Delta = 0$; in other words, $\phi(x)$ is continuous at $x = 0$.

We note that the regularity condition (1.2, iv) is in a sense superfluous since it serves merely to determine $\phi(x)$ uniquely at every point of the interval (0, 1). With this remark the proof of our theorem is complete.

The analogy between Hausdorff integral and matrix transformations is far from being perfect. In fact, one of the most significant of the Hausdorff matrix transformations, the Euler transformation, is trivial with respect to the Hausdorff integral transformation theory. We recall that the mass function $\phi(x)$ for the Euler method of summation has a single discontinuity at the point $x = \delta$, $0 < \delta < 1$, and otherwise remains constant ([5] or [6]). Thus, for this method of summation, we have

$$v(x) = \int_0^x u(y)\, d\phi(y/x) = \int_0^1 u(xy)\, d\phi(y) = u(\delta x).$$

Since $\lim_{x\to\infty} u(\delta x)$ exists or fails to exist with $\lim_{x\to\infty} u(x)$ it follows that the Euler method of summation associated with integral transformations is identical with convergence. This observation is of importance with respect to the main theorem of the next section.

3. The Silverman-Schmidt integral equations[3]

The principal results of this section involve the restricted class of Hausdorff mass functions satisfying the following conditions:

(3.1)
(i) $\phi(x)$ is of bounded variation on the interval $0 \leqq x \leqq 1$,
(ii) $\phi(x)$ is continuous on the interval $(0, 1)$ except possibly at $x = 1$,
(iii) $\phi(0) = 0$,
(iv) $\phi(1) = 1$.

First of all we shall establish the following theorem.

THEOREM 2. *If the functions $\phi_a(x)$, $\phi_b(x)$, $\phi_c(x)$ correspond to the transformations A, B, C respectively, and if $C = AB$, then $\phi_c(x)$ is given in terms of $\phi_a(x)$ and $\phi_b(x)$ by the Silverman-Schmidt equations*:

(3.2)
(i) $$\phi_c(x) = \int_0^1 \phi_a(x/t)\, d\phi_b(t),$$
(ii) $$\phi_c(x) = \int_0^1 \phi_b(x/t)\, d\phi_a(t).$$

If $\phi_a(x)$ and $\phi_b(x)$ satisfy the conditions (3.1), *then so also does $\phi_c(x)$. All transformations involving mass functions of the type* (3.1) *are permutable and hence consistent.*

---

[3] Evidently the equations (3.2) were first derived by Silverman [1], but not as Stieltjes integral equations. Later, the equations (3.2) were used by R. Schmidt [7] in an investigation completely unrelated to the inclusion problems of the theory of summability. Hille and Tamarkin [3] established the relationship of these equations to inclusion problems in the domain of Hausdorff matrix transformations, and in this connection they have been studied recently by Garabedian, Hille, and Wall [4].

The transformation A, with the mass function $\phi_a(x)$ satisfying the conditions (3.1), is written in the form

$$v(x) = \int_0^x u(y)\, d\phi_a(y/x) = \int_0^1 u(xs)\, d\phi_a(s).$$

In the last integral let us remove a possible discontinuity of $\phi_a(x)$ at $x = 1$. We get

$$v(x) = u(x)[\phi_a(1) - \phi_a(1-0)] + \int_0^1 u(xs)\, d\theta_a(s),$$

where $\theta_a(x)$ is identical with $\phi_a(x)$ in the interval $0 \leqq x < 1$. Notice that $\theta_a(x)$ is continuous on the interval $0 \leqq x \leqq 1$ and $\theta_a(1) = \phi_a(1-0)$. The transformation B is obtained by replacing a by b in the preceding discussion.

We wish to investigate the form of the transformation C:

$$\begin{aligned} w(x) &= \int_0^x u(s)\, d\phi_c(s/x) = \int_0^1 u(xs)\, d\phi_c(s) \\ &= u(x)[\phi_c(1) - \phi_c(1-0)] + \int_0^1 u(xs)\, d\theta_c(s), \end{aligned} \tag{3.3}$$

where $C = BA$, A taking u into v and B taking v into w.

We have

$$w(x) = v(x)[1 - \phi_b(1-0)] + \int_0^1 v(xt)\, d\theta_b(t),$$

and hence

$$w(x) = \beta\left[\alpha u(x) + \int_0^1 u(xs)\, d\theta_a(s)\right] + \int_0^1 \left[\alpha u(xt) + \int_0^1 u(xst)\, d\theta_a(s)\right] d\theta_b(t).$$

where $\alpha = 1 - \phi_a(1-0)$, $\beta = 1 - \phi_b(1-0)$. Then

$$\begin{aligned} w(x) = \alpha\beta u(x) &+ \alpha \int_0^1 u(xs)\, d\theta_b(s) + \beta \int_0^1 u(xs)\, d\theta_a(s) \\ &+ \int_0^1 \int_0^1 u(xst)\, d\theta_a(s)\, d\theta_b(t). \end{aligned}$$

If we set $st = y$ in the last integral we obtain

$$\begin{aligned} w(x) = \alpha\beta u(x) &+ \alpha \int_0^1 u(xs)\, d\theta_b(s) + \beta \int_0^1 u(xs)\, d\theta_a(s) \\ &+ \int_0^1 \int_0^t u(xy)\, d\theta_a(y/t)\, d\theta_b(t). \end{aligned}$$

Now, we find it convenient to define $\theta_a(x) = \phi_a(1-0)$, $\theta_b(x) = \phi_b(1-0)$, $x \geqq 1$. Then, we may write

$$w(x) = \alpha\beta u(x) + \alpha \int_0^1 u(xs)\, d\theta_b(s) + \beta \int_0^1 u(xs)\, d\theta_b(s) + \int_0^1 \int_0^1 u(xy)\, d\theta_a(y/t)\, d\theta_b(t).$$

The last integral is of a type studied by Bray [8, Th. 5, p. 183] relative to a change in the order of integration. In the present situation the hypotheses of Bray's theorem are fulfilled with a comfortable margin of safety. Accordingly, we have

$$w(x) = \alpha\beta u(x) + \alpha \int_0^1 u(xs)\, d\theta_b(s) + \beta \int_0^1 u(xs)\, d\theta_a(s) + \int_0^1 u(xy)\, d_y \int_0^1 \theta_a(y/t)\, d\theta_b(t). \tag{3.4}$$

Comparing (3.4) with (3.3) we have

$$\theta_c(s) = \alpha\theta_b(s) + \beta\theta_a(s) + \int_0^1 \theta_a(s/t)\, d\theta_b(t), \tag{3.5}$$

or

$$\theta_c(s) = \alpha\theta_b(s) + \beta\theta_a(s) + \theta_a(1)\theta_b(s) + \int^1 \theta_a(s/t)\, d\theta_b(t),$$

or

$$\theta_c(s) = \theta_b(s) + \beta\theta_a(s) + \int_s^1 \theta_a(s/t)\, d\theta_b(t). \tag{3.6}$$

Observe also that from (3.5) we can write

$$\phi_c(s) = \int_0^1 \phi_a(s/t)\, d\phi_b(t) = \phi_b(s) + \int_s^1 \phi_a(s/t)\, d\phi_b(t). \tag{3.7}$$

We now proceed to show that $\phi_c(x)$ satisfies the conditions (3.1). From (3.7) we have $\phi_c(1) = 1$. Now, in (3.6) we set $s = ty$ to obtain

$$\theta_c(s) = \theta_b(s) + \beta\theta_a(s) - \int_s^1 \theta_a(y)\, d\theta_b(s/y).$$

It follows now, from a theorem of Bray [8, Th. 3, p. 180], that $\theta_c(x)$ is continuous on the interval $0 \leqq x \leqq 1$ and hence that $\theta_c(0) = 0$. It results *a fortiori* that the conditions (ii) and (iii) of (3.1) are fulfilled. Finally, if we consider (3.6) and use another result of Bray [8, Th. 4, p. 181], it follows that $\theta_c(s)$ and hence $\phi_c(s)$ is of bounded variation on the interval (0, 1).

Using (3.7) and integrating by parts we obtain

$$\phi_c(s) = \phi_a(s) - \int_s^1 \phi_b(t)\, d\phi_a(s/t).$$

If we set $s = ty$ we get

$$\phi_c(s) = \phi_a(s) + \int^1 \phi_b(s/y)\, d\phi_a(y). \tag{3.8}$$

This equation is of the same form as (3.7) except that ϕ_a and ϕ_b have been interchanged. Hence, $AB = BA$, whence the transformations A and B are permutable. This completes the proof of Theorem 2.

We are now in a position to state and prove the main theorem of this paper.

THEOREM 3. *Let $\phi_a(x)$ and $\phi_b(x)$ satisfy the conditions* (3.1). *If there exists a solution $\phi_c(x)$ of the type* (3.1) *which satisfies either of the Silverman-Schmidt equations*:

$$\begin{aligned} \phi_a(x) &= \int_0^1 \phi_c(x/t)\, d\phi_b(t), \\ \phi_a(x) &= \int_0^1 \phi_b(x/t)\, d\phi_c(t), \end{aligned} \tag{3.9}$$

then $(H, \phi_a(x)) \supset (H, \phi_b(x))$.

We wish to find sufficient conditions on the mass functions $\phi_a(x)$ and $\phi_b(x)$ in order that $(H, \phi_a(x)) \supset (H, \phi_b(x))$. Suppose that $\phi_a(x)$ and $\phi_b(x)$ satisfy the conditions (3.1). Let $u(x)$ be any function transformed by $(H, \phi_a(x))$ and $(H, \phi_b(x))$ into $v(x)$ and $w(x)$ respectively. We seek conditions under which $\lim_{x\to\infty} w(x) = \xi$ implies $\lim_{x\to\infty} v(x) = \xi$. Symbolically we have

$$\begin{aligned} w(x) &= B\{u(x)\} \to \xi, \\ v(x) &= A\{u(x)\}. \end{aligned}$$

Suppose that C is any transformation with an associated mass function of the type (3.1). Since C is regular we have

$$C\{w(x)\} = CB\{u(x)\} \to \xi.$$

If there exists a transformation C such that $A = CB$, then

$$\begin{aligned} \xi &= \lim_{x\to\infty} C\{w(x)\} = \lim_{x\to\infty} CB\,\{u(x)\} \\ &= \lim_{x\to\infty} A\{u(x)\} = \lim_{x\to\infty} v(x). \end{aligned}$$

In other words, if there exists a solution $\phi_c(x)$ of the type (3.1) satisfying either of the Silverman-Schmidt equations, we have $(H, \phi_a(x)) \supset (H, \phi_b(x))$.

4. Implications of Theorem 3

In the field of Hausdorff matrix transformations it is known that $[H, \phi_a(x)] \supset [H, \phi_b(x)]$, $\phi_a(x)$ and $\phi_b(x)$ being mass functions of the type (1.2), if and only if

Now, we find it convenient to define $\theta_a(x) = \phi_a(1-0)$, $\theta_b(x) = \phi_b(1-0)$, $x \geqq 1$. Then, we may write

$$w(x) = \alpha\beta u(x) + \alpha \int_0^1 u(xs)\, d\theta_b(s) + \beta \int_0^1 u(xs)\, d\theta_b(s) + \int_0^1 \int_0^1 u(xy)\, d\theta_a(y/t)\, d\theta_b(t).$$

The last integral is of a type studied by Bray [8, Th. 5, p. 183] relative to a change in the order of integration. In the present situation the hypotheses of Bray's theorem are fulfilled with a comfortable margin of safety. Accordingly, we have

$$w(x) = \alpha\beta u(x) + \alpha \int_0^1 u(xs)\, d\theta_b(s) + \beta \int_0^1 u(xs)\, d\theta_a(s) + \int_0^1 u(xy)\, d_y \int_0^1 \theta_a(y/t)\, d\theta_b(t). \tag{3.4}$$

Comparing (3.4) with (3.3) we have

$$\theta_c(s) = \alpha\theta_b(s) + \beta\theta_a(s) + \int_0^1 \theta_a(s/t)\, d\theta_b(t), \tag{3.5}$$

or

$$\theta_c(s) = \alpha\theta_b(s) + \beta\theta_a(s) + \theta_a(1)\theta_b(s) + \int_0^1 \theta_a(s/t)\, d\theta_b(t),$$

or

$$\theta_c(s) = \theta_b(s) + \beta\theta_a(s) + \int_s^1 \theta_a(s/t)\, d\theta_b(t). \tag{3.6}$$

Observe also that from (3.5) we can write

$$\phi_c(s) = \int_0^1 \phi_a(s/t)\, d\phi_b(t) = \phi_b(s) + \int_s^1 \phi_a(s/t)\, d\phi_b(t). \tag{3.7}$$

We now proceed to show that $\phi_c(x)$ satisfies the conditions (3.1). From (3.7) we have $\phi_c(1) = 1$. Now, in (3.6) we set $s = ty$ to obtain

$$\theta_c(s) = \theta_b(s) + \beta\theta_a(s) - \int_s^1 \theta_a(y)\, d\theta_b(s/y).$$

It follows now, from a theorem of Bray [8, Th. 3, p. 180], that $\theta_c(x)$ is continuous on the interval $0 \leqq x \leqq 1$ and hence that $\theta_c(0) = 0$. It results *a fortiori* that the conditions (ii) and (iii) of (3.1) are fulfilled. Finally, if we consider (3.6) and use another result of Bray [8, Th. 4, p. 181], it follows that $\theta_c(s)$ and hence $\phi_c(s)$ is of bounded variation on the interval (0, 1).

Using (3.7) and integrating by parts we obtain

$$\phi_c(s) = \phi_a(s) - \int_s^1 \phi_b(t)\, d\phi_a(s/t).$$

If we set $s = ty$ we get

$$\phi_c(s) = \phi_a(s) + \int^1 \phi_b(s/y)\, d\phi_a(y). \tag{3.8}$$

This equation is of the same form as (3.7) except that ϕ_a and ϕ_b have been interchanged. Hence, $AB = BA$, whence the transformations A and B are permutable. This completes the proof of Theorem 2.

We are now in a position to state and prove the main theorem of this paper.

THEOREM 3. *Let $\phi_a(x)$ and $\phi_b(x)$ satisfy the conditions* (3.1). *If there exists a solution $\phi_c(x)$ of the type* (3.1) *which satisfies either of the Silverman-Schmidt equations*:

$$\phi_a(x) = \int_0^1 \phi_c(x/t)\, d\phi_b(t), \qquad \phi_a(x) = \int_0^1 \phi_b(x/t)\, d\phi_c(t), \tag{3.9}$$

then $(H, \phi_a(x)) \supset (H, \phi_b(x))$.

We wish to find sufficient conditions on the mass functions $\phi_a(x)$ and $\phi_b(x)$ in order that $(H, \phi_a(x)) \supset (H, \phi_b(x))$. Suppose that $\phi_a(x)$ and $\phi_b(x)$ satisfy the conditions (3.1). Let $u(x)$ be any function transformed by $(H, \phi_a(x))$ and $(H, \phi_b(x))$ into $v(x)$ and $w(x)$ respectively. We seek conditions under which $\lim_{x\to\infty} w(x) = \xi$ implies $\lim_{x\to\infty} v(x) = \xi$. Symbolically we have

$$w(x) = B\{u(x)\} \to \xi,$$

$$v(x) = A\{u(x)\}.$$

Suppose that C is any transformation with an associated mass function of the type (3.1). Since C is regular we have

$$C\{w(x)\} = CB\{u(x)\} \to \xi.$$

If there exists a transformation C such that $A = CB$, then

$$\xi = \lim_{x\to\infty} C\{w(x)\} = \lim_{x\to\infty} CB\,\{u(x)\}$$

$$= \lim_{x\to\infty} A\{u(x)\} = \lim_{x\to\infty} v(x).$$

In other words, if there exists a solution $\phi_c(x)$ of the type (3.1) satisfying either of the Silverman-Schmidt equations, we have $(H, \phi_a(x)) \supset (H, \phi_b(x))$.

4. Implications of Theorem 3

In the field of Hausdorff matrix transformations it is known that $[H, \phi_a(x)] \supset [H, \phi_b(x)]$, $\phi_a(x)$ and $\phi_b(x)$ being mass functions of the type (1.2), if and only if

there exists a mass function $\phi_c(x)$ of the type (1.2) satisfying the Silverman-Schmidt equations (3.9), for all except at most a countable set of values of x in the interval $0 < x < 1$ [4]. Thus, the Silverman-Schmidt equations serve as a connecting link between the theories of Hausdorff matrix and integral transformations. This relationship has already been discussed at considerable length by this writer in a paper already referred to [2]. It will suffice to note here that all of the inclusion relationships among Hausdorff matrix transformations involving mass functions of the type (1.4) were shown to be valid in the field of Hausdorff integral transformations. Since mass functions of the type (3.1) embrace a far wider class of Hausdorff mass functions than the absolutely regular mass functions of Silverman, Theorem 3 extends considerably the class of known inclusion relationships among Hausdorff integral transformations. We provide an example in illustration of the last statement.

Let us consider briefly the mass functions

$$\phi_1(x) = \begin{cases} 2x, & 0 \leqq x \leqq 1/2, \\ 1, & 1/2 < x \leqq 1, \end{cases}$$

$$\phi_2(x) = x, \qquad 0 \leqq x \leqq 1.$$

We observe that $\phi_1(x)$ and $\phi_2(x)$ are mass functions of the type (3.1). In the field of Hausdorff matrix transformations it is easily proved that $[H, \phi_1(x)] \supset [H, \phi_2(x)]$. It follows that there exists a solution of the type (3.1) of the corresponding Silverman-Schmidt integral equations. We conclude finally that $(H, \phi_1(x)) \supset (H, \phi_2(x))$, and thus obtain a hitherto unknown inclusion relationship between Hausdorff integral transformations.

In this writer's opinion there is little hope of obtaining Theorem 3 for a much more general class of mass functions than those included in (3.1). There is some evidence available in support of this statement. Let us first observe that it is necessary to allow for a possible discontinuity of $\phi(x)$ at $x = 1$, since it is known that if $(H, \phi_a(x)) \approx (H, \phi_b(x))$ or $[H, \phi_a(x)] \approx [H, \phi_b(x)]$ then a solution $\phi_c(x)$ of the Silverman-Schmidt equations will have a discontinuity at $x = 1$ [2]. If we allow for additional discontinuities of $\phi(x)$ in the interval $0 < x \leqq 1$, in special cases, at least, we should be led to contradictions. For example, let $\phi_a(x)$ and $\phi_b(x)$ be the mass functions associated with Euler methods of summation whose associated *moment sequences* [6] are $\{\delta_a^n\}$ and $\{\delta_b^n\}$, $0 < \delta_a < \delta_b < 1$. We have $[H, \phi_a(x)] \supset [H, \phi_b(x)]$, while (*vide supra*) $(H, \phi_a(x)) \approx (H, \phi_b(x))$, both methods of summation being equivalent to convergence. Consequently, the Silverman-Schmidt equations would lead to contradictory results in this case. Evidently, all mass functions of step-function character are associated with integral transformations equivalent to each other and to convergence. Whatever improvements in generality it would be possible to make in Theorem 2 would involve the tedious and difficult matter of reproving the theorems of Bray used in this paper for mass functions of a very special character.

NORTHWESTERN UNIVERSITY

Bibliography

1. Silverman, L. L., *The equivalence of certain regular transformations*, Trans. Am. Math. Soc., vol. 26 (1924), pp. 101–112.
2. Garabedian, H. L., *A class of linear integral transformations*, Am. Jour. of Math., vol. 64 (1942), pp. 208–214.
3. Hille, E. and Tamarkin, J. D., *Questions of relative inclusion in the domain of Hausdorff means*, Proc. Nat. Acad. Sci., vol. 19 (1933), pp. 573–577.
4. Garabedian, H. L., Hille, E., and Wall, H. S., *Formulations of the Hausdorff inclusion problem*, Duke Math. Jour., vol. 8 (1941), pp. 193–213.
5. Hausdorff, F., *Summationsmethoden und Momentfolgen, I and II*, Math. Zeit., vol. 9 (1921), pp. 74–109, 280–299.
6. Garabedian, H. L., *Hausdorff matrices*, Am. Math. Monthly, vol. 46 (1939), pp. 390–410.
7. Schmidt, R., *Über divergente Folgen und lineare Mittelbildungen*, Math. Zeits., vol. 22 (1925), pp. 89–152.
8. Bray, H. E., *Elementary properties of the Stieltjes integral*, these Annals, vol. 20 (1919), pp. 177–186.

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

THE PROBLEM OF DIFFUSION OF WAVES

BY J. HADAMARD

(Received March 26, 1942)

To the memory of MYRON MATHISON, whose premature death is a cruel loss to Science, I dedicate this treatment of the problem which he has solved so beautifully.

1

The various forms of Huygens' principle concern (as studied heretofore) phenomena governed by linear partial differential equations of the second order[1]

$$\text{(E)} \qquad F(u) = A^{kl}\frac{\partial^2 u}{\partial x^k\,\partial x^l} + \cdots = \begin{cases} 0 & \text{(homogeneous equation)} \\ f(x) & \text{(inhomogeneous equation),} \end{cases}$$

where the terms replaced by dots contain the unknown u itself or its first derivatives, the A's, the other coefficients, and f being given functions of the independent variables x. I have previously considered three of them,[2] two of which —the "major premise" and what can be called the "conclusion," if Huygens' argument is considered as a syllogism—are general properties of such a class of phenomena. On the contrary, the "minor premise" is only true for quite special equations of the type (E): it means that when a disturbance originally located within a determinate finite region of space, propagates by waves and reaches any given point outside that region after a certain time, no effect persists *after* the passing of the wave. This is what Mathisson described by saying that these waves are *pure*.

Now, waves are not pure for any equation in an odd number of independent variables, (i.e. for wave motions in spaces with an even number of dimensions), nor are they pure when the number of variables is two. On the other hand, as is well known, the classical equation of spherical waves

$$\text{(E}_0\text{)} \qquad \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} - \frac{\partial^2 u}{\partial z^2} = 0$$

gives rise to pure waves, and this is also the case for its analogues in 6, 8, $\cdots$, variables. The question is whether these cases are the only ones. Of course,

[1] We shall not use, at least in the present paper, the methods of the absolute differential calculus. However, we shall speak of contravariant and covariant vectors, the components being denoted by superscripts in the first case, subscripts in the second.

As usual in the absolute differential calculus, we omit, when no confusion can arise, the summation signs referring to repeated superscripts and subscripts.

[2] See my lecture on the subject in the Bulletin de la Société Mathématique de France, vol. LII, 1925.

two equations of the type (E) are to be considered as not essentially different if they can be deduced from each other by: 1) any point transformation

(a) $$\bar{x}^h = (x^1, x^2, \cdots, x^m) \qquad (h = 1, 2, \cdots, m)$$

on the x's; 2) the multiplication of both sides of (E) by any (non-vanishing) factor, say

(b) $$\frac{1}{\lambda} F(u) = \begin{cases} 0, \\ \dfrac{1}{\lambda} f(x), \end{cases} \qquad \lambda = \lambda(x) = e^{\mu(x)} ;$$

3) multiplying the unknown by any non-vanishing factor, say

(c) $$F(\lambda u) = \begin{cases} 0, \\ f. \end{cases}$$

The combination of the last two, viz. replacing (E) by

(bc) $$F_\lambda(u) = \frac{1}{\lambda} F(\lambda u) = e^{-\mu} F(ue^{\mu}) = \begin{cases} 0, \\ e^{-\mu} f(x) \end{cases}$$

permits an easier treatment, on account of the fact that it does not change the terms of the second order and, therefore, does not alter the "characteristic form"

$$\mathbf{A}(p_1, p_2, \cdots, p_m ; x^1, x^2, \cdots, x^m) = A^{kl} p_k p_l$$

nor the corresponding "metric form"

$$\mathfrak{A}(dx^1, dx^2, \cdots, dx^m; x^1, x^2, \cdots, x^m) = A_{kl} dx^k dx^l .$$

This combination is the only one which we shall deal with in this first paper.

The relations between the above two forms, which are reciprocal ones, are well known. To obtain A_{kl}, we have to divide by Δ($\Delta = 1/D$ being the discriminant of $\mathbf{A}$; D, the discriminant of $\mathfrak{A}$) the coefficient of A^{kl} in Δ. We have

$$A^{kh} A_{kj} = \begin{cases} 0 & (h \neq j), \\ 1 & (h = j). \end{cases}$$

The differentials dx, or, to deal with finite quantities, the m derivatives $\dot{x}$ are to be taken as the components of a contravariant vector, the covariant components of which are the p's if we have

$$\dot{x}^h = A^{hk} p_k ,$$

or the equivalent equations

$$p_k = A_{hk} \dot{x}^h .$$

This also implies

$$\mathbf{A}(p) = \mathfrak{A}(\dot{x})$$

so that the left and right members, one being the characteristic form and the other the metric form, represent one and the same form expressed once in terms of the covariant components, then in terms of the contravariant ones.

2

Now Mathisson has succeeded in giving the above question an affirmative answer, i.e. in proving that no other simply hyperbolic[3] linear partial equation in four independent variables generates pure waves, except (E_0) and those equations which are not essentially different from it. Mathisson divides his proof into two stages:

1) In the first place, he considers the case in which the coefficients A of the terms of the second order are constants;

2) In a further discussion he treats the general case.

The first part alone has been published[4] and it will be the only one we shall deal with at present.

3

In the special case of constant A's, we shall not have to use transformation (a), or (b) or (c) alone, but only (bc).

Furthermore, if the A's are constant, it will always be allowed (the equation being of the simply hyperbolic type) to suppose that they have the same values as in (E_0), viz.

$$(3.1) \qquad A^{hk} = 0 \ (h \neq k), \qquad A^{11} = A^{22} = A^{33} = -1, \qquad A^{44} = 1.$$

However, we shall begin by not even using the first assumption; let us study the effect of (bc) on the *general* equation (E). As for the second assumption (3.1), it will be useful not to use it until the last step of the calculation.[5]

In the case of variable A's, it is convenient to write (E) or, rather, the adjoint equation[6]

$$G(v) = 0$$

in the form[7]

$$(\mathfrak{E}) \qquad \Delta_2 v + B^h \frac{\partial v}{\partial x^h} + Cv = 0,$$

$$\Delta_2 v = \frac{1}{\sqrt{D}} \frac{\partial}{\partial x^h} \left(\sqrt{D}\, A^{hk} \frac{\partial v}{\partial x^k} \right) \text{(second differential parameter of Lamé-Beltrami)}.$$

---

[3] This means that the characteristic form **A** or the equivalent metric form $\mathfrak{A}$ consists of one positive and three negative squares: what we previously called the *normal* hyperbolic case. We adopt Hilbert's and Courant's terminology.

[4] Acta Mathematica, Vol. LXXI, 1939, pp. 249–282.

[5] The disadvantage of immediately assuming (3.1) lies in the asymmetry of the equations between the suffixes 1, 2, 3 on one hand; 4 on the other.

[6] In passing from (E) to its adjoint, (b) and (c) are permuted, so that (bc) remains unchanged but for the change of sign of μ.

[7] Σ's are omitted, as remarked in the beginning.

[8] See, e.g., Darboux, *Leçons sur la théorie des surfaces*, vol. III.

It is known[8] that

$$\Delta_2(\lambda v) = \lambda\Delta_2 v + 2\Delta_1(\lambda, v) + v\Delta_2\lambda,$$

$\Delta_1(\lambda, v) = A^{hk}(\partial v/\partial x^h)\cdot(\partial\lambda/\partial x^k)$ (first parameter of Lamé-Beltrami for two functions), and that

$$e^{-\mu}\Delta_2(e^{\mu}) = \Delta_2\mu + \Delta_1\mu,$$

$\Delta_1\mu = \Delta_1(\mu, \mu) = A^{hk}(\partial\mu/\partial x^h)(\partial\mu/\partial x^k)$, (first parameter of Lamé-Beltrami for one function) so that

$$e^{-\mu}g(ve^{\mu}) = \Delta_2 v + 2\bar{B}^h\frac{\partial v}{\partial x^h} + \bar{C}v, \tag{3.2}$$

$$\bar{B}^h = B^h + 2A^{hk}\frac{\partial\mu}{\partial x^k}, \tag{3.3}$$

$$\bar{C} = e^{-\mu}G(e^{\mu}) = \Delta_2\mu + \Delta_1\mu + B^h\frac{\partial\mu}{\partial x^h} + C. \tag{3.4}$$

The transformation formula (3.3) on the B's suggests substituting, in place of the quantities B^h considered as contravariant components of a vector, the corresponding covariant components

$$B_k = A_{hk}B^h$$

with respect to which the transformation formulae are simply

$$\bar{B}_k = B_k + 2\frac{\partial\mu}{\partial x^k}. \tag{3.3'}$$

We see that the differences $\bar{B}_k - B_h$ are the partial derivatives of one and the same function μ: therefore, the expressions

$$H_{hk} = \frac{\partial B_h}{\partial x^k} - \frac{\partial B_k}{\partial x^h} \tag{3.5}$$

are invariants with respect to (bc).

4

The answer to our question rests, of course, on the integration of (E) itself: more precisely, on the solution of the corresponding Cauchy problem. This integration is obtained by the use of a proper solution v of the adjoint equation, by means of which the value of the unknown u in Cauchy's problem concerning (E) is directly given by (in our case, quadruple, triple and double) integrals extended over parts of the region where the right-hand member f is defined and of the variety to which Cauchy's data refers.[9]

Instead of using the exact elementary solution, one can also get to a result

[9] See our *Lectures on Cauchy's problem*, Cambridge-New Haven, 1923 or the French edition, Paris 1932, especially Liv. IV.

by introducing an approximate solution—what Hilbert calls a "parametrix": then the answer to Cauchy's problem is not directly obtained, but is reduced to an integral equation of the Fredholm (or rather Volterra) type. This is what Mathisson does. A first difficulty which he has to overcome consists, then, in showing that the necessary and sufficient condition in order that waves be pure is that the parametrix thus introduced be an exact solution, after which he has to find the cases where this condition is satisfied.

Now, it seemed to me that this roundabout use of an integral equation by introduction of a special parametrix could be avoided, since the exact elementary solution can actually be constructed; and indeed, it seems that the proof becomes simpler this way. Let us recall how the elementary solution can be calculated.[10]

$a(a_1, a_2, \cdots, a_m)$ being an arbitrarily given point, chosen as the singular point or "pole" of the solution in question, let us consider the various geodesics issuing from a relative to the metric form $\mathfrak{A}$ and denote by $\Gamma = \Gamma(x^1, \cdots, x^m, a^1, \cdots, a^m)$ the square of the geodesic distance betweeh a and any point x: a first term (the number of independent variables still being assumed to be 4) will be V_0/Γ, with

$$V_0 = \exp\left\{-\frac{1}{2}\int_a^x \left(\frac{M}{2} - m\right)\frac{ds}{s}\right\}, \tag{4.1}$$

where

$$M = G(\Gamma) - C\Gamma = \Delta_2\Gamma + B^h\frac{\partial\Gamma}{\partial x^h}, \tag{4.1'}$$

the integral being taken along the geodesic ax and s denoting the independent variable in Hamilton's equations

$$(\mathcal{H}) \qquad \dot{x}^h = \frac{dx^h}{ds} = \frac{1}{2}\frac{\partial \mathbf{A}}{\partial p_h}, \qquad \dot{p}_h = \frac{dp_h}{ds} = -\frac{1}{2}\frac{\partial \mathbf{A}}{\partial x^h}$$

for that geodesic ($\mathbf{A}$ is precisely the characteristic form, considered in Section 1).

In general, $G(V_0)$ will be different from zero; then, in order to obtain the elementary solution, the first term V_0/Γ must be completed by a term in $\log \Gamma$. The latter vanishes if, and only if

$$G(V_0) = 0 \tag{4.2}$$

whenever the point x belongs to the characteristic conoid which has its vertex at a. Now, the general theory shows[11] that *this is the necessary and sufficient condition in order that the waves be pure.*

5

Following the same general principle as Mathisson—and we shall be able to apply it even more thoroughly than he himself has done—, we shall write our equation in a simplified form with respect to the group of the transformations

[10] *Lectures on Cauchy's problem,* Liv. II. chap. III.

[11] *Loc. cit.,* Liv. IV, chap. I.

(bc). Now, this is obtained at once by the consideration of V_0 : for, by an arbitrary transformation (bc), V_0 is changed into V_0/λ, as is seen by the formulae (4.1), (4.1′) and (5.2) (see below): therefore, by taking $\lambda = V_0$, we get, in one and only one way, an equation for which

$$V_0 = 1, \tag{5.1}$$

a condition which, it must be remembered, *depends on the choice of the pole a.*

The above reduction condition means that the integrand in (4.1) must be identically zero in x. In the special case of constant A's, treated in the present memoir, Γ is simply the quadratic form $\Gamma(\xi) = A_{hk}\xi^h\xi^k$; the symbol Δ_2 reduces to $A^{hk}\,\partial^2/\partial x^h\partial x^k$, so that, remembering (1), $\Delta_2\Gamma - 2m$ vanishes identically. As for $B^h\,\partial\Gamma/\partial x^h$, the derivatives of Γ being $\partial\Gamma/\partial x^h = 2P_h = 2sp_h$, it can be written, in any case, on account of $(\mathcal{H})$,

$$B^h\frac{\partial\Gamma}{\partial x^h} = 2sA^{hk}B_k p_k = 2sB_h\dot{x}^h, \tag{5.2}$$

$\dot{x}^h$ standing for a derivative taken along the geodesic ax. When the coefficients of the terms of second order are constant, the geodesics are straight lines along which each x varies linearly, so that $s\dot{x}^h = \xi^h = x^h - a^h$. In this case, therefore, $V_0 = 1$ gives us

$$B_h\xi^h = 0. \tag{5.3}$$

We shall assume that the B's are regular in the neighborhood of a, and therefore have a Taylor's expansion

$$B_h = (B_h)_a + \xi^k\left(\frac{\partial B_h}{\partial x^k}\right)_a + \cdots.$$

Substituting in (5.3), we see that, in the reduced form of the equation, we must have,[12] at a,

$$(5.4)_a \qquad \left\{\begin{array}{ll} (5.4)_a & \bar{B}_h = 0, \\ (5.4')_a & \dfrac{\partial\bar{B}_h}{\partial x^k} + \dfrac{\partial\bar{B}_k}{\partial x^h} = 0, \\ (5.4'')_a & \dfrac{\partial^2\bar{B}_h}{\partial x^k\,\partial x^l} + \dfrac{\partial^2\bar{B}_k}{\partial x^h\,\partial x^l} + \dfrac{\partial^2\bar{B}_l}{\partial x^h\,\partial x^k} = 0, \\ \cdots\cdots & \end{array}\right\} \quad (h, k, l, \cdots = 1, 2, 3, 4).$$

If, instead of taking λ equal to V_0 , we had only chosen it tangent to V_0 at a up to certain order, we should not have all the relations $(5.4)_a$, but only a certain (arbitrarily large) number of them. It is remarkable that Mathisson seems to have been led to write these conditions a priori, while we deduce them from (5.1), which seems to be their true origin.

[12] Throughout the following, we shall add the suffix a after the number of every relation which is proved only for $x = a$, in order to distinguish them from those which are identities in the x's.

6

In this section, we consider the general equation (&), not assuming the A's to be constant.

The successive relations $(5.4)_a$ yield the successive derivatives of $\bar{B}_h$ or, more precisely, the numerical values which they assume at a, which we shall denote by $\mu_h = (\partial\mu/\partial x^h)_a$, $\mu_{hk} = (\partial^2\mu/\partial x^h \partial x^k)_a$, $\cdots$. The transformation formulae will be, *at the point a*,

$$(6.1)_a \qquad \bar{B}_h = B_h + 2\mu_h,$$

$$(6.1')_a \qquad \frac{\partial \bar{B}_h}{\partial x^k} = \frac{\partial B_h}{\partial x^k} + 2\mu_{hk},$$

$$\cdots\cdots\cdots\cdots\cdots$$

together with the corresponding formula for $\bar{C}$ obtained from (3.4), viz.

$$(6.2)_a \qquad \bar{C} = C + B^h \mu_h + A^{hk} \mu_h \mu_k + \frac{1}{2} \frac{\partial \log |D|}{\partial x^h} A^{hk} \mu_k + \frac{\partial A^{hk}}{\partial x^h} \mu_k + A^{hk} \mu_{hk}.$$

While the set of transformations (bc) constitutes an infinite group, formulae such as the above define finite ones, since μ_h, μ_{hk}, $\cdots$ are nothing but numerical parameters: $(6.1)_a$ are the equations of a group in m parameters; $(6.1)_a$, $(6.1')_a$ and $(6.2)_a$, the equations of a group in $m + \frac{1}{2}m(m+1)$ parameters; and so on. In order to obtain the reduced values, we must take

$$(6.3)_a \qquad \begin{cases} 2\mu_h = -B_h, \\ 2\mu_{hk} = -\frac{1}{2}\left(\frac{\partial B_h}{\partial x^k} + \frac{\partial B_k}{\partial x^h}\right), \\ 2\mu_{hkl} = -\frac{1}{3}\left(\frac{\partial^2 B_h}{\partial x^k \partial x^l} + \frac{\partial^2 B_k}{\partial x^l \partial x^h} + \frac{\partial^2 B_l}{\partial x^h \partial x^k}\right), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

Now, having these values of μ_h, μ_{hk}, $\cdots$, we see, in the first place, that the reduction is made in one and only one way, so that *every equation of the class under consideration has one determinate reduced homologue with respect to the group* (bc).

The values assumed at a by the coefficients of this reduced homologue and of their derivatives at a can be calculated in terms of the original coefficients and their derivatives, by means of (6.1), (6.1') $\cdots$ and (6.2); we find, still at a

$$(6.4)_a \qquad \begin{cases} \bar{B}_h = 0, \\ \frac{\partial \bar{B}_h}{\partial x^k} = \frac{1}{2}\left(\frac{\partial B_h}{\partial x^k} - \frac{\partial B_k}{\partial x^h}\right) = \frac{1}{2} H_{hk}, \\ \frac{\partial^2 \bar{B}_h}{\partial x^k \partial x^l} = \frac{1}{3}\left(\frac{\partial H_{hk}}{\partial x^l} + \frac{\partial H_{hl}}{\partial x^k}\right), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

then, remembering $(6.2)_a$, where, on count of $(6.3)_a$, there is a reduction between the second and the third term, we have

$$(6.5)\qquad \bar{C} = C - \frac{1}{4} B^h B_h - \frac{1}{4} \frac{\partial \log |D|}{\partial x^h} B^h - \frac{1}{2} \frac{\partial A^{hk}}{\partial x^h} B_k - \frac{1}{4} A^{hk} \left(\frac{\partial B_h}{\partial x^k} + \frac{\partial B_k}{\partial x^h} \right)$$

where[13] the last two terms can be replaced by $-\frac{1}{2}\partial B^h/\partial x^h$.

All the quantities in the right-hand members are invariants[14] *of $G(v)$ with respect to the group* (bc). We already knew those of the first line in $(6.4)_a$, and the other ones which contain only the B's can be deduced from them by differentiation, so that they do not bring us anything new. But such is not the case as concerns the last one (6.5).

The calculations in the present section being independent[14] of the hypothesis that the A's be constant, the quantities (3.5), (6.5) are invariants, with respect to (bc), for every equation such as $(\mathfrak{E})$.

From now on, we come back to the case of constant A's in which the invariant (6.5) is simplified by the vanishing of the terms in $\partial A^{hk}/\partial x^h$ or in $\partial \log |D|/\partial x^h$.

7

If we write $(\mathfrak{E})$ in its reduced form with respect to the pole a, so that equations $(5.4)_a$ are satisfied, the condition $G(V_0) = 0$ is nothing else than $\bar{C} = 0$. This must happen not only at the point $x = a$, but along the whole surface of the characteristic conoid—which, in the present case, is not distinct from the characteristic cone of vertex a. In the first place we must have, at a,

$$(7.1)\qquad \bar{C} = 0$$

and this property belongs to the original (i.e. not reduced) equation—a consequence of the fact that (6.5) is an invariant. Since this must be true whatever the point a may be, we see that (7.1) *must be an identity*: we can replace C everywhere, by its value obtained from it and (6.5).

Furthermore, at a, we must have

$$(7.2)_a\qquad \frac{\partial(\bar{C})}{\partial x^h} = 0,$$

$$(7.3)_a\qquad \frac{\partial^2(\bar{C})}{\partial x^h \partial x^k} = qA^{hk},$$

where q is a factor common to all the terms $(7.3)_a$; and further successive relations expressing the fact that the function $(\bar{C})$ (assumed to be regular) is divisible by Γ.

---

[13] This time, we do not write the suffix a, as explained in the next section.

[14] For variable A's, the equations $(6.3)_a$, $(6.4)_a$ $\cdots$ no longer agree with (5.1) (with the exception of the first line in (6.4)); but this is not necessary in the present section, it being essential only that the conditions define one determinate reduced homologue. We could have replaced the right hand members in (5.4) by arbitrarily given quantities Q_h, Q_{hk}, $\cdots$, modifying the μ_h, μ_{hk}, $\cdots$ accordingly. It is easy to verify that we should have obtained, in that way, the same invariants with only the addition of terms containing exclusively A's and the Q's.

The derivatives	must be replaced in a derivative of $\bar{C}$, by	must be replaced in a derivative of $(\bar{C})$, by
$\dfrac{\partial \mu_k}{\partial x^h}$	$\dfrac{\partial \mu_k}{\partial x^h} = -\dfrac{1}{2}\dfrac{\partial B_k}{\partial x^k}$	$\left(\dfrac{\partial \mu_k}{\partial x^h}\right) = -\dfrac{1}{4}\left(\dfrac{\partial B_k}{\partial x^h} + \dfrac{\partial B_h}{\partial x^k}\right)$
$\dfrac{\partial^2 \mu_k}{\partial x^h\, \partial x^j}$	$\dfrac{\partial^2 \mu_k}{\partial x^h\, \partial x^j} = -\dfrac{1}{2}\dfrac{\partial^2 B_k}{\partial x^h\, \partial x^j}$	$\left(\dfrac{\partial^2 \mu_k}{\partial x^h\, \partial x^j}\right) = -\dfrac{1}{6}\left(\dfrac{\partial^2 B_k}{\partial x^h\, \partial x^j} + \dfrac{\partial^2 B_j}{\partial x^h\, \partial x^k} + \dfrac{\partial^2 B_h}{\partial x^j\, \partial x^k}\right)$
.....	.....	.....
$\dfrac{\partial \mu_{kl}}{\partial x^h}$	$\dfrac{\partial \mu_{kl}}{\partial x^h} = -\dfrac{1}{4}\left(\dfrac{\partial^2 B_k}{\partial x^h\, \partial x^l} + \dfrac{\partial^2 B_l}{\partial x^h\, \partial x^k}\right)$	$\left(\dfrac{\partial \mu_{kl}}{\partial x^h}\right) = -\dfrac{1}{6}\left(\dfrac{\partial^2 B_h}{\partial x^k\, \partial x^l} + \dfrac{\partial^2 B_k}{\partial x^h\, \partial x^l} + \dfrac{\partial^2 B_l}{\partial x^h\, \partial x^k}\right)$
$\dfrac{\partial^2 \mu_{kl}}{\partial x^h\, \partial x^j}$	$\left(\dfrac{\partial^2 \mu_{kl}}{\partial x^h\, \partial x^j}\right) = -\dfrac{1}{4}\left(\dfrac{\partial^3 B_k}{\partial x^h\, \partial x^j\, \partial x^l} + \dfrac{\partial^3 B_l}{\partial x^h\, \partial x^j\, \partial x^k}\right)$	$\left(\dfrac{\partial^2 \mu_{kl}}{\partial x^h\, \partial x^j}\right) = -\dfrac{1}{8}\left(\dfrac{\partial^3 B_k}{\partial x^h\, \partial x^j\, \partial x^l} + \dfrac{\partial^3 B_l}{\partial x^h\, \partial x^j\, \partial x^k} + \dfrac{\partial^3 B_h}{\partial x^k\, \partial x^l\, \partial x^j} + \dfrac{\partial^3 B_j}{\partial x^k\, \partial x^l\, \partial x^h}\right),$
.....	.....	.....

Since, in (7.1), $\bar{C}$ is obtained by identifying x with a, and, thus (7.1) becomes an identity, $\bar{C}$ and $(\bar{C})$ must be distinguished from each other: in (7.1), $\bar{C}$ is the value of (6.2) obtained by replacing, at any point a (or x), μ and its derivatives by their numerical values corresponding to reduction at the *same* point, while, in $(7.2)_a$, $(7.3)_a$, $\cdots$. $(\bar{C})$, at any point x, means the value of (3.4) when μ is constructed corresponding to reduction at a, viz. $\mu = -\log V_0(x; a)$, so that its successive derivatives, at a, are given by the formulae $(6.3)_a$.

Both quantities $\bar{C}$ and $(\bar{C})$ are derived from (6.2) by substituting $-\frac{1}{2}B_h$ for μ_h and $-\frac{1}{4}(\partial B_h/\partial x^k + \partial B_k/\partial x^h)$ for μ_{hk}, so that, if not differentiated, their values are equal; another treatment is needed for their derivatives. (Cf. Table.)

Note that the corresponding values of one and the same derivative disagree only by some permutations of suffixes, so that their difference can always be expressed as a combination of the invariants H defined by (3.5).

8

Let us apply this, in the first place, to $\partial\bar{C}/\partial x^h$ and $\partial(\bar{C})/\partial x^h$, both of which we must equate to zero. We have

$$\frac{\partial\bar{C}}{\partial x^h} = \frac{\partial C}{\partial x^h} + \frac{\partial B^k}{\partial x^h}\mu_k + (B^k + 2A^{kl}\mu_l)\frac{\partial\mu_k}{\partial x^h} + A^{kl}\frac{\partial\mu^{kl}}{\partial x^h}. \tag{8.1}$$

The coefficient of $\partial\mu_k/\partial x^h$ vanishes, as μ_l is defined by the first formula $(6.3)_a$. Moreover, μ_k is the same in both expressions which we have to consider. Therefore, substracting them from each other in order to eliminate the derivative of C, the result, simplified by interchanges of k with l in the coefficients of A^{kl}, reduces to

$$\begin{aligned}\frac{1}{6}H'_h &= A^{kl}\left[\left(\frac{\partial\mu_{kl}}{\partial x^h}\right) - \frac{\partial\mu_{kl}}{\partial x^h}\right] \\ &= \frac{1}{2}A^{kl}\left[\frac{\partial^2 B_k}{\partial x^h\,\partial x^l} - \frac{1}{3}\left(2\frac{\partial^2 B_k}{\partial x^h\,\partial x^l} + \frac{\partial^2 B_h}{\partial x^k\,\partial x^l}\right)\right] = \frac{1}{6}A^{kl}\frac{\partial H_{kh}}{\partial x^l} = 0,\end{aligned} \tag{8.2}$$

which, for the same reason as (7.1), *are identities*.

9

These conditions are always satisfied if the H_{hk} are constant: especially, therefore, if the B's are linear functions of the x's. We shall now show a remarkable consequence of this, that if complex coefficients were admitted,[15] waves could be pure for equations essentially distinct from (E_0). If we do not mind introducing imaginaries, nothing prevents us from replacing, in the terms of the second order, the assumption (1) by

$$A^{hk} = 0 \quad h \neq k, \qquad A^{hh} = 1 \qquad (h = 1, 2, 3, 4). \tag{9.1}$$

[15] Complex values could not be considered for the A's, as the distinction between elliptic and hyperbolic cases would lose its meaning; but, theoretically, a similar objection does not hold with regard to the coefficients B.

Let us take, for each B_h, a linear polynomial

$$B_h(x) = b_h(x) + \beta_h = b_h(\xi) + B_h(a),$$
$$b_h(\xi) = \alpha_{hj}\xi^j$$

being the homogeneous part. On account of (3.2′), we can alter these expressions by the elements of any exact differential. Therefore, decomposing the matrix $|| \alpha_{hj} ||$ into a symmetric one plus an antisymmetric one, the former can be cancelled and we can assume

$$\alpha_{hj} = -\alpha_{jh}, \tag{9.2}$$

and in particular

$$\alpha_{hh} = 0. \tag{9.2′}$$

Then $\sum \xi^h b_h(\xi)$ vanishes and we have

$$V_0 = \exp\{-\tfrac{1}{2}\xi^h B_h(a)\} = e^w. \tag{9.3}$$

On the other hand, C is given by (6.5), in which the last term disappears on account of (9.2′). Thus, substituting V_0 in G, we get

$$\frac{1}{V_0} G(V_0) = e^{-w} G(e^w) = \tfrac{1}{4}\{\sum [B_h(a)]^2 - 2\sum B_h(a)B_h(x) + C\}$$
$$= \tfrac{1}{4}\{\sum [B_h(a)]^2 - 2\sum B_h(a)B_h(x) + \sum [B_h(x)]^2\}$$
$$= \tfrac{1}{4}\sum [b_h(\xi)]^2.$$

This must be zero on the conoid which has its vertex at a, so that the above quadratic form must be proportional to $\sum_0^3 (\xi^h)^2$, and we are led to the problem of *expressing the sum* $\sum (\xi^h)^2$ *as a sum of squares of linear forms, the coefficients of which form an antisymmetric matrix.*

As is easily seen, the most general way of obtaining this is to complete (9.2) into

$$\alpha_{hj} = -\alpha_{jh} = \epsilon\alpha_{kl} = -\epsilon\alpha_{lk}, \qquad \epsilon = \pm 1, \tag{9.4}$$

where $hjkl$ is any even (or alternate) permutation of the suffixes 1, 2, 3, 4.

In other words, we have only to take $\alpha = 0$ in the well-known identity

$$(\alpha^2 + \beta^2 + \gamma^2 + \delta^2)(x^2 + y^2 + z^2 + t^2)$$
$$= (\alpha x + \beta y + \gamma z + \delta t)^2 + (-\beta x + \alpha y + \gamma z + \delta t)^2$$
$$+ (-\gamma x - \delta y + \alpha z + \beta t)^2 + (-\delta x + \gamma y - \beta z + \alpha t)^2$$

which gives the transformation of the product of two sums of four squares into a sum of four squares. A simple form of the result, to which the most general one can be reduced by an orthogonal change of the variables (for instance, an

orthogonal change of x and z combined with an orthogonal change on x and t) is the equation

$$F(u) = \Delta u + 2\left(y\frac{\partial u}{\partial x} - x\frac{\partial u}{\partial y} + t\frac{\partial u}{\partial z} - z\frac{\partial u}{\partial t}\right) + u(x^2 + y^2 + z^2 + t^2) = 0$$

which, for any given a, b, c, d, admits of the solution

$$u = \frac{V_0}{(x-a)^2 + (y-b)^2 + (z-c)^2 + (t-d)^2} - w$$

$$= \frac{e^{ay - bx + ct - dz}}{(x-a)^2 + (y-b)^2 + (z-c)^2 + (t-d)^2} - w,$$

w being any holomorphic solution of $F(u) = V_0$.

Changing t into it, this would give a hyperbolic equation corresponding to pure waves. But we have reached this only by introduction of imaginaries. On the contrary, we shall now see that no such solution can exist in the real domain.

10

Let us treat $(7.3)_a$ as we did for $(7.2)_a$. We have

$$\frac{\partial^2 \bar{C}}{\partial x^h \partial x^j} = \frac{\partial^2 C}{\partial x^h \partial x^j} + \frac{\partial^2 B^k}{\partial x^h \partial x^j}\mu_k + \left(\frac{\partial B^k}{\partial x^h}\frac{\partial \mu_k}{\partial x^j} + \frac{\partial B^k}{\partial x^j}\frac{\partial \mu_k}{\partial x^h}\right)$$

$$+ 2A^{kl}\frac{\partial \mu_l}{\partial x^h}\frac{\partial \mu_k}{\partial x^j} + (B^k + 2A^{kl}\mu_l)\frac{\partial^2 \mu_k}{\partial x^h \partial x^j} + A^{kl}\frac{\partial^2 \mu_{kl}}{\partial x^h \partial x^j}$$

As in the treatment of $(7.2)_a$, we remark that the coefficient of $\partial^2\mu_k/\partial x^h \partial x^j$ vanishes and the second term is the same in $\partial^2\bar{C}/\partial x^h \partial x^j$ and in $\partial^2(\bar{C})/\partial x^h \partial x^j$. As for the terms in $\partial^2\mu_{kl}/\partial x^h \partial x^j$, they give, in $\partial^2(\bar{C})/\partial x^h \partial x^j - \partial^2\bar{C}/\partial x^h \partial x^j$, the result

$$A^{kl}\left[\frac{1}{2}\frac{\partial^3 B_k}{\partial x^h \partial x^j \partial x^l} - \frac{1}{8}\left(2\frac{\partial^3 B_k}{\partial x^h \partial x^j \partial x^l} + \frac{\partial^3 B_j}{\partial x^h \partial x^k \partial x^l} + \frac{\partial^3 B_h}{\partial x^j \partial x^k \partial x^l}\right)\right]$$

$$= \frac{1}{8}A^{kl}\left(\frac{\partial^2 H_{kh}}{\partial x^j \partial x^l} + \frac{\partial^2 H_{kj}}{\partial x^h \partial x^l}\right) = \frac{1}{8}\left(\frac{\partial H'_h}{\partial x^j} + \frac{\partial H'_j}{\partial x^h}\right)$$

which vanishes on account of (8.2). Moreover, we have

$$\left(\frac{\partial \mu_k}{\partial x^j}\right) - \frac{\partial \mu_k}{\partial x^j} = \frac{1}{4}H_{kj}, \qquad \left(\frac{\partial \mu_k}{\partial x^h}\right) - \frac{\partial \mu_k}{\partial x^h} = \frac{1}{4}H_{kh}$$

so that we can write for the difference $\dfrac{\partial^2(\bar{C})}{\partial x^h \partial x^j} - \dfrac{\partial^2 \bar{C}}{\partial x^h \partial x^j}$,

$$\frac{\partial^2(\bar{C})}{\partial x^h \partial x^j} - \frac{\partial^2 \bar{C}}{\partial x^h \partial x^j} = qA^{hj} = \frac{A^{kl}}{2}\left[\frac{\partial B_k}{\partial x^j}H_{lh} + \frac{\partial B_l}{\partial x^h}H_{kh}\right.$$

(10.1)

$$\left. + 2\left(-\frac{\partial B_l}{\partial x^h} + \frac{1}{2}H_{lh}\right)\left(-\frac{\partial B_k}{\partial x^j} + \frac{1}{2}H_{kj}\right) - 2\frac{\partial B_l}{\partial x^h}\frac{\partial B_k}{\partial x^j}\right] = \frac{1}{4}A^{kl}H_{lh}H_{kj},$$

the same as that which Mathisson obtains by his method.[16] As he shows, this immediately gives the desired solution: using, this time, the special choice (3.1) of the coefficients A, we only need to take, in (10.1), $h = j = 4$, then $h = j = 1$, thus obtaining

$$q = -H_{14}^2 - H_{24}^2 - H_{34}^2,$$
$$-q = H_{41}^2 - H_{21}^2 - H_{31}^2,$$

and by adding these two equations we see that we must have $H_{24} = H_{34} = H_{21} = H_{31} = 0$; and similarly for the other quantities H. Then, since every H vanishes, $B_h\,dx^h$ must be an exact differential $-d\mu$ and there exists a transformation (bc) which reduces all the coefficients B to zero, after which C must also be zero, on account of (7.1) and (6.5).

NEW YORK

[16] It is remarkable that Mathisson's deductions and ours follow a quite analogous order, successively obtaining (8.2) and then (10.1), although his parametrix is infinite along a parallel to the x^0-axis, while our v_0 is a regular function.

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

A PROOF OF THE FUNDAMENTAL THEOREM ON THE DENSITY OF SUMS OF SETS OF POSITIVE INTEGERS*

BY HENRY B. MANN

(Received January 7, 1942; revised March 16, 1942)

Let $A(B, C)$ be sets of positive integers. Form A^0, B^0 by adjoining 0 to A and B respectively. Let $A(n)$ be the number of positive integers in A that are $\leqq n$. The greatest lower bound of the quotients $A(n)/n$ is called the density of A. Let C^0 consist of all integers of the form $a + b (a \in A^0, b \in B^0)$.

Let α be the density of A, β the density of B, γ the density of C, then we shall prove

$$\gamma \geqq \alpha + \beta \text{ or } = 1. \tag{1}$$

This inequality has been conjectured by E. Landau, I. Schur and A. Khintchine.

Approximations to (1) have been obtained by the following authors:

E. Landau:[1] $\gamma \geqq \alpha + \beta - \alpha\beta$.

A. Besicovitch:[2] If α^* is the lower bound of the quotients $A(n)/(n+1)$ then $\gamma \geqq \alpha^* + \beta$ or $= 1$.

I. Schur:[3] $\gamma \geqq \alpha/(1-\beta)$ or $= 1$.

I. Schur:[4] $\gamma \geqq \frac{1}{2}\alpha + \frac{1}{2}(\alpha^2 + 4\beta^2)^{\frac{1}{2}}$ or $= 1$.

A. Brauer:[5] $\gamma \geqq 9/10(\alpha + \beta)$ or $= 1$.

An important partial result was obtained by A. Khintchine:[6] If α_i is the density of A_i and $\alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_n$ and $C^0 = \{a_1 + a_2 + \cdots + a_n\} (a_i \in A_i^0)$ then $\gamma \geqq n\alpha_1$ or $= 1$. In this paper (1) will be proved completely.

Stripped of its transcendental content (1) states that

$$\frac{C(n)}{n} = 1 \quad \text{or} \quad \geqq \min. \frac{A(l)}{l} + \frac{B(m)}{m}, \qquad 1 \leqq m \leqq n, \qquad 1 \leqq l \leqq n.$$

We propose to prove the following sharper theorem.

FUNDAMENTAL THEOREM: *Let $A(B, C)$ be sets of positive integers. Let $A(n)$, $B(n)$, $C(n)$ be the numbers of the integers $1, 2, \cdots, n$ that are in A, B, C*

* Presented to the American Mathematical Society Feb. 28, 1942. The author's interest in this problem was aroused through Dr. A. T. Brauer's lectures on additive theory of numbers at New York University.

[1] *Die Goldbachsche Vermutung und der Schnirelmannsche Satz.* Gottinger Nachrichten, Math. Phys. Klasse (1930), pp. 255–276.

[2] *On the density of the sum of two sequences of integers.* Jour. London Math. Soc., Vol. 10 (1935), pp. 246–248.

[3] *Über den Begriff der Dichte in der additiven Zahlentheorie.* Sitzungsber. der preuss. Akad. der Wiss., Math. Phys. Klasse, (1936), pp. 269–297.

[4] L.c. footnote 3.

[5] These Annals, Vol. 42, (1941), pp. 959–988.

[6] *Zur additiven Zahlentheorie.* Matematiceski Sbornik, Vol. 39, (1932) pp. 27–34.

respectively. Let $C^0 = \{a + b\}(a \in A^0,\ b \in B^0)$ *where* A^0, B^0 *consist of the numbers of* A *and* B *respectively and the number* 0. *Then*

$$(2) \qquad \frac{C(n)}{n} = 1 \quad or \quad \geqq \min. \frac{A(m) + B(m)}{m}, \qquad 1 \leqq m \leqq n, \qquad m \notin C.$$

Let $n_1 < n_2 \cdots$ be the numbers of $\bar{C}$ (i.e. numbers not in C); then $C(n_r) = n_r - r$. (2) is an immediate consequence of the inequality

$$(3) \qquad \frac{C(n_r)}{n_r} \geqq \min. \frac{A(n_q) + B(n_q)}{n_q}, \qquad q \leqq r.$$

We divide the numbers $\leqq n_r$ into three sets: numbers of B, numbers of the form $n_r - a(a \in A^0)$, numbers of L_r where L_r denotes the set of all positive numbers $\leqq n_r$ that are neither in B nor of the form $n_r - a(a \in A^0)$. Denote by l_r the number of integers in L_r. These three sets are disjoint. Hence

$$(4) \qquad n_r = B(n_r) + A(n_r) + 1 + l_r.$$

(3) follows from (4) for $r = 1$. In proving (3) by induction we may assume

$$\frac{C(n_s)}{n_s} > \frac{C(n_r)}{n_r} \quad \text{or} \quad rn_s > sn_r \quad \text{for} \quad s = 1, \cdots, r-1.$$

Our statement then is that

$$(5) \qquad \frac{A(n_s) + B(n_s)}{n_s} > \frac{C(n_r)}{n_r} \quad \text{for} \quad s = 1, \cdots, r-1$$

implies

$$(6) \qquad C(n_r) \geqq A(n_r) + B(n_r),$$

or using (4)

$$\frac{1 + l_s}{n_s} < \frac{r}{n_r} \quad \text{for} \quad s = 1 \cdots r-1 \quad \text{implies} \quad 1 + l_r \geqq r.$$

We have to prove the following lemma: *If* $rn_s > sn_r$, $rn_s > (1 + l_s)n_r$ *for* $s = 1, 2, \cdots, r-1$, *then* $l_r \geqq r - 1$.

From now on r is fixed. We therefore write $n_r = N$, $L_r = L$, $l_r = l$. s, t range from 1 to $r - 1$. Put $d_s = N - n_s$ then $n_s - d_t = n_t - d_s$.

Construction of numbers of L. To prove the lemma we have to construct $r - 1$ numbers b' in L, i.e. numbers not in B and not of the form $N - a(a \in A^0)$. To satisfy both conditions we construct numbers b' satisfying the equations

$$b' = n_s - a = N - \bar{a} \qquad (a \in A^0,\ \bar{a} \notin A^0).$$

The condition $\bar{a}$ not $\in A^0$ is satisfied by putting $\bar{a} = n_t - b$ or

$$b' = b + d_t = n_s - a.$$

Thus we arrive at the following recursive definition of numbers $e_1, e_2, \cdots$, and sets $B^0, B^1, B^2, \cdots$, and $T_1, T_2, \cdots$, starting with B^0.

DEFINITION: a) *We choose* e_{i+1} *as an element in* $B^0 \smile B^1 \smile \cdots \smile B^i$.

b) t *lies in* T_{i+1} *if there is an* $a \in A^0$ *such that*

$$a + e_{i+1} + d_t = n_s \tag{7}$$

and neither s *nor* t *in* $T_1 \smile T_2 \smile \cdots \smile T_i$.

c) $B^{i+1} = \{e_{i+1} + d_t\}_{t \in T_{i+1}}$.

Equation (7) is equivalent to

$$a + e_{i+1} + d_s = n_t, \tag{7'}$$

hence s as well as t lies in T_{i+1}. We may therefore write

$$b' = e_{i+1} + d_t = n_s - a \in B^{i+1}. \tag{8}$$

The following propositions follow easily from the definition:

1. $e_j \in B^0 \smile B^1 \smile \cdots \smile B^{j-1} \quad (j \geqq 1)$.
2. *No element of* T_j *lies in* $T_1 \smile T_2 \smile \cdots \smile T_{j-1}$.
3. $B^j = \{e_j + d_t\}_{t \in T_j}$.

COROLLARY: B^j *contains as many elements as* T_j. *All elements of* B^j *are greater than* e_j.

4. *An element* b' *of* B^j *may be written as* $n_s - a (s \in T_j, a \in A^0)$. *Also* n_s *can be decomposed into* $a + b' (a \in A^0, b' \in B^j)$.
5. *If* $a + b' = n_s$, $(a \in A^0, b' \in B^0 \smile B^1 \smile \cdots B^j)$ *then* $s \in T_1 \smile \smile \cdots \smile T_j$.

PROOF: Proposition 5 is true for $j = 0$. We assume that it is true for $j = i$ and prove it for $j = i + 1$. If $b' \in B^{i+1}$ then $b' = e_{i+1} + d_t$, $t \in T_{i+1}$, hence $t \notin T_1 \smile T_2 \smile \cdots \smile T_i$, therefore $a + e_{i+1} + d_t = n_s$, $a + e_{i+1} + d_s = n_t$. If $s \notin T_1 \smile T_2 \smile \cdots \smile T_i$, then $s \in T_{i+1}$ according to definition of T_{i+1}.

6. *No element of* B^j $(j \geqq 1)$ *lies in* $B^0 \smile B^1 \smile \cdots B^{j-1}$ *nor is of the form* $N - a$ $(a \in A^0)$.

PROOF: An element b_j of B^j is of the form $n_s - a (s \in T_j, a \in A^0)$. But the relation

$$n_s = a + b_j, \qquad s \notin T_1 \smile T_2 \smile \cdots \smile T_{j-1}, \qquad b_j \in B^0 \smile B^1 \smile \cdots \smile B^{j-1}$$

contradicts 5. Each element b_j of B^j is of the form $e_j + d_s$ $(s \in T_j)$: the equation $e_j + d_s = N - a$, $(a \in A^0)$ or $a + e_j = n_s$ contradicts 5 because of

$$s \notin T_1 \smile T_2 \smile \cdots \smile T_{j-1}, \qquad e_j \in B^0 \smile B^1 \smile \cdots \smile B^{j-1}.$$

In order to arrive at a uniquely determined construction we shall choose e_{i+1} as the *least* number for which T_{i+1} is not empty. The final construction is then defined as follows:

a) t *lies in* $T_{i+1}\{e\}$ *if there is an* $a \in A^0$ *such that*

$$a + e + d_t = n_s$$

and neither t *nor* s *in* $T_1 \smile \cdots \smile T_i$.

b) e_{i+1} *is the smallest* e *in* $B^0 \smile B^1 \smile \cdots \smile B^i$ *for which* $T_{i+1}\{e\}$ *is not empty.*

c) $T_{i+1} = T_{i+1}\{e_{i+1}\}$, $B^{i+1} = \{e_{i+1} + d_t\}_{t \in T_{i+1}}$.

The process is to be continued until $T_{J+1}\{e\}$ is empty for every $e \in B^0 \smile B^1 \smile \cdots \smile B^J$.

We are then able to prove the following propositions:

7. $T_{J+1}\{e\}$ *is empty for every* $e \in B^0 \cup B^1 \cup \cdots \cup B^J$.

8. *If* $e \in B^0 \cup B^1 \cup \cdots \cup B^{j-1}$ *and* $T_j\{e\}$ *is not vacuous then* $e \geqq e_j$.

9. $e_1 < e_2 < \cdots < e_J$.

PROOF: $e_j = e_{j+1}$ is impossible because $T_{j+1}\{e_j\}$ is empty. $e_j < e_{j+1}$ means therefore that there is no $e < e_j$ for which $e \in B^0 \cup \cdots \cup B^j$ and $T_{j+1}\{e\}$ is not empty. If there were, then $T_j\{e\}$ would not be empty. Either $e \in B^0 \cup \cdots \cup B^{j-1}$, and then $e \geqq e_j$ according to 8; or $e \in B^j$, and then $e > e_j$ according to the corollary to 3.

10. *All elements of* $B^j \cup \cdots \cup B^J$ *are greater than* e_j $(1 \leqq j \leqq J)$. *This follows from the corollary to* 3 *and from* 9.

11. *Let* s *be the least index not in* $T_1 \cup T_2 \cup \cdots \cup T_J$. *Let there be* q *indices* $(r-1, r-2, \cdots, r-q)$ *such that* $d_t \leqq n_s$. *Then*

$$l_s \geqq q.$$

PROOF: $n_s - d_t$ $(d_t \leqq n_s)$ is either in C^0 and then $= a + b'$ $(a \in A^0, b' \in B^0)$ or $= n_u$, $(u \in T_1 \cup \cdots \cup T_J)$ and hence $= a + b'$ $(a \in A^0, b' \in B^0 \cup \cdots \cup B^J)$. Thus in either case

$$n_s - d_t = a + b' \ (a \in A^0, b' \in B^0 \cup \cdots \cup B^J).$$

If t is not in $T_1 \cup \cdots \cup T_J$ then $T_{J+1}\{b'\}$ is not empty which contradicts 7.

Let $t \in T_j$, $e_j + d_t = b_j \in B^j$. We assert $e_j + d_t \leqq n_s$. If $e_j + d_t > n_s$, then

$$e_j > n_s - d_t = a + b' \geqq b'.$$

Thus b' cannot lie in $B^j \cup \cdots \cup B^J$ (see Prop. 10). Hence $b' \in B^0 \cup \cdots \cup B^{j-1}$. The equation

$$a + b' + d_t = n_s$$

with $t \notin T_1 \cup \cdots \cup T_{j-1}$, $s \notin T_1 \cup \cdots \cup T_{j-1}$ shows that $T_j\{b'\}$ is not vacuous and $b' < e_j$ is impossible by Prop. 8.

b_j is not in B nor of the form $n_s - a$ $(a \in A^0)$ (see 5). Hence it is in L_s. Because of 6 we obtain q distinct such numbers b_j corresponding to the q values of t. Therefore

$$l_s \geqq q.$$

We can prove now our lemma. Suppose that

$$rn_i > iN, \qquad rn_i > (1 + l_i)N \quad \text{for } i = 1, 2, \cdots, r-1.$$

We shall show that $T_1 \cup \cdots \cup T_J$ contain all indices $i \leqq r-1$. Because of 6 and the corollary to 3 our lemma is an immediate consequence of this fact. Suppose s is the least index not in $T_1 \cup \cdots \cup T_J$. Let q be defined as in Proposition 11. Then

$$rn_s > (1 + l_s)N \geqq (q+1)N, \qquad rn_{(r-q-1)} > (r-q-1)N.$$

Adding these two inequalities we obtain

$$n_s + n_{(r-q-1)} > N \qquad \text{or} \qquad n_s > d_{(r-q-1)}.$$

But then we would have $q+1$ indices $t=(r-1, r-2, \cdots, r-q-1)$ for which $d_t \leqq n_s$ contrary to the significance of q. This proves our lemma.

Another result of a different nature can be obtained by the construction method used in the proof of the fundamental theorem. We propose to prove

THEOREM II. *If* $C^0=\{a+b\}(a \in A^0, b \in B^0)$ *and if* α^* *is the G.L.B. of the quotients* $A(m)/(m+1)$ *and if* n *is not in* C. *Then*

$$C(n) \geqq \alpha^* n + B(n). \tag{9}$$

From equation (4) we have for $r=1$.

$$C(n_1) \geqq \alpha^* n_1 + B(n_1) + (\alpha - \alpha^*)n_1 .$$

We propose to prove by induction

$$C(n_i) \geqq \alpha^* n_i + B(n_i) + (\alpha - \alpha^*)n_1 , \quad i = 1, 2, \cdots \tag{9'}$$

We distinguish two cases: 1). $n_r - n_{(r-1)} > 1$; 2). $n_r - n_{(r-1)} = 1$.

PROOF FOR CASE 1: The integers $> n_{(r-1)}$ and $\leqq n_r$ are either in B or of the form $n_r - a$ $(a \in A^0, 0 \leqq a < n_r - n_{(r-1)})$ or in neither of these two sets. Hence

$$\begin{aligned} n_r - n_{(r-1)} &\geqq A(n_r - n_{(r-1)} - 1) + B(n_r) - B(n_{(r-1)}) + 1 \\ &\geqq \alpha^*(n_r - n_{(r-1)}) + B(n_r) - B(n_{(r-1)}) + 1. \end{aligned}$$

By induction we have

$$C(n_{(r-1)}) = n_{(r-1)} - (r-1) \geqq \alpha^* n_{(r-1)} + B(n_{(r-1)}) + (\alpha - \alpha^*)n_1 .$$

Adding both inequalities we obtain (9′) for $i=r$.

PROOF FOR CASE 2: In this case we have $d_{(r-1)} = 1$, $n_1 - d_{(r-1)} = a + b$ $(a \in A^0, b \in B^0)$. We form $T_1\{b\}$ and B^1. Let B^* contain the numbers of B^0 and B^1. Put $C^* = \{a + b'\}$ $(a \in A^0, b' \in B^*)$. n_r will then be the j^{th} gap $j < r$ in C^* and we have by induction

$$C^*(n_r) \geqq \alpha^* n_r + B^*(n_r) + (\alpha - \alpha^*)n_1 .$$

But by proposition 5 we have

$$C^*(n_r) - C(n_r) = B^*(n_r) - B(n_r).$$

Subtracting this equation we obtain (9′) for $i=r$.

It is not possible to substitute α for α^* in (9) as shown by the following example

$$\begin{aligned} A = B &= \{1, 2, 6, 7, 8, 12, \cdots\}, \\ C &= \{1, 2, 3, 4, 6, 7, 8, 9, 10, 12, \cdots\}, \\ \alpha &= \tfrac{2}{5}, \qquad B(11) = 5, \qquad C(11) = 9. \end{aligned}$$

If $C^0 = \{a + b\}$ $(a \in A^0, b \in B)$ and if $1 \in B$ then it can be shown exactly by the same method that $C(n) \geqq \alpha^* n + B(n)$ if n is not in C.

COLUMBIA UNIVERSITY

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

A REMARK ON S. KAKUTANI'S CHARACTERIZATION OF (L)-SPACES[1]

BY M. F. SMILEY

(Received March 9, 1942)

The purpose of this note is to show that we may rephrase the condition (IX) of Kakutani in a way which indicates a close relation between the norm in an (AL)-space and the metric of a metric lattice.[2] This will also permit a slight weakening of assumptions in §§3–5.

We show first that the condition

$$\text{(IX')} \qquad \| x - y \| = \| x \vee y - x \wedge y \|$$

is equivalent to (IX) under (I)–(IV), (VI)–(VII).

PROOF. (IX) *implies* (IX'): By Lemma 1.5, $x = x_+ - x_-$ and $x_+ \wedge x_- = 0$; consequently $\| x \| = \| x_+ - x_- \| = \| x_+ + x_- \|$ by (IX). Since $-x \vee 0 = -(x \wedge 0)$, we have the equation

$$\| x \| = \| x \vee 0 + (-x \vee 0) \| = \| x \vee 0 - x \wedge 0 \|.$$

Replacing x by $x - y$ and applying Lemma 1.2 yields (IX').

(IX') *implies* (IX): If $x \wedge y = 0$, then $x \vee y = x + y$ by Lemma 1.3. Hence, if $x \wedge y = 0$, the condition (IX') yields $\| x - y \| = \| x \vee y \| = \| x + y \|$.

We may now prove that (I)–(IV), (VI)–(IX) imply the following condition.

$$(1) \qquad \| a \vee x - a \vee y \| + \| a \wedge x - a \wedge y \| = \| x - y \|.$$

PROOF.[3] By (IX') and Lemma 1.6 the left side of (1) is $\| a \vee x \vee y - a \vee (x \wedge y) \| + \| a \wedge (x \vee y) - a \wedge x \wedge y \|$. Condition (VIII) reduces this to the expression

$$\| a \vee x \vee y - a \vee (x \wedge y) + a \wedge (x \vee y) - a \wedge x \wedge y \|.$$

Using Lemma 1.3 we obtain the identity (1) at once.

The identity (1) shows that the assumption (V) is not needed in §§3–5. We may even show that the triangle inequality for the norm is a consequence of (I)–(IV), (VI)–(IX) by a simple paraphrase of the proof of Theorem 3.10 of G. Birkhoff.[4]

---

[1] Presented to the American Mathematical Society, April 3, 1942. Our title refers to Kakutani's paper *Concrete representation of abstract (L)-spaces and the mean ergodic theorem*, these Annals, vol. 42 (1941), pp. 523–537. All references are to this paper unless the contrary is explicitly stated.

[2] Cf. G. Birkhoff, *Lattice Theory*, Amer. Math. Soc. Col. Pub., vol. 25 (1940), p. 41.

[3] *Ibid.*, p. 42.

[4] *Ibid.*, p. 42.

NOTE ADDED IN PROOF. G. Birkhoff has remarked to the author that $m[a] \equiv \| a_+ \| - \| a_- \|$ is a sharply positive modular functional in an (L)-space. This result depends only on conditions (I)–(IV), (VI)–(VIII). For, if $a \geqq b$, then $m[a] - m[b] = \| a - b \|$ by (VIII) and Lemma 1.2. Since $\| a \| > 0$ unless $a = 0$, this shows that $m[a]$ is sharply positive. Applying this result again we find that $m[a] - m[a \wedge b] = \| a - a \wedge b \| = \| a \vee b - b \| = m[a \vee b] - m[b]$. Thus $m[a]$ is also a modular functional. Our condition (IX′) then merely requires that the metric $\delta(a, b)$ which arises from $m[a]$ should coincide with $\| a - b \|$. When this is true, the identity (1) is a consequence of (in fact, equivalent to) the distributive law.

LEHIGH UNIVERSITY.

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

GENERALIZED SURFACES IN THE CALCULUS OF VARIATIONS. II

BY L. C. YOUNG

(Received December 21, 1941)

MEAN SURFACES AND THE THEORY OF THE PROBLEM $\int\int f(x, y, p, q)\,dx\,dy = \text{Min.}$

1. Introduction

In this second note, machinery is produced for studying extensions of the classical necessary conditions applicable both in the ordinary and in the generalized minimum problem for surface integrals. We employ it to derive necessary and sufficient conditions in the case of an integrand $f(x, y, p, q)$ in which z does not occur explicitly. Our methods follow naturally from the idea of generalized surface developed in the first note,[1] but are no longer restricted to continuous integrands $f(x, y, z, p, q)$. We show, in fact, that for a much wider class of these integrands the minimum in the problem considered is unaltered if we admit, in addition to ordinary surfaces, only a special kind of generalized surface which we term "mean surface." The ideas connected with the notion of a mean surface are closely akin to a method developed by Haar[2] in the special case of a regular problem. They lead to a generalization of an inequality, due to Steiner,[3] valid when the integrand has the form $f(x, y, p, q)$; and thence to equations characterizing a Lipschitzian surface for which our minimum is attained.

2. Mean surfaces

Let z_1 and z_2 denote the tracks, and $M_1(g)$ and $M_2(g)$ the averages at x, y, of two generalized surfaces S_1 and S_2 or, in particular, of two ordinary surfaces. We define the *mean* of S_1 and S_2 as the generalized surface S with the track $z = \frac{1}{2}z_1 + \frac{1}{2}z_2$, and with the average $M(g) = \frac{1}{2}M_1(g) + \frac{1}{2}M_2(g)$ at the point x, y. A generalized surface which is the mean of two *ordinary* surfaces will be termed shortly a *mean surface*.

(2.1) *Any mean surface is expressible as the mean of two ordinary surfaces whose tracks differ by at most* ϵ.

To prove this, it is clearly sufficient to show that if S is the mean of two ordinary surfaces S_1 and S_2 whose tracks z_1 and z_2 differ by at most $2a$, then S is also expressible as the mean of a second pair of ordinary surfaces whose tracks differ by at most a. Let E, E', E'' denote the sets of x, y in which, respectively,

$$|z_1 - z_2| < a, \qquad z_1 - z_2 \leqq -a, \qquad z_1 - z_2 \geqq a;$$

[1] Young [12].
[2] Haar [3].
[3] Steiner [10], p. 298.

and let $\bar{z}_1$, $\bar{z}_2$ denote the pair of functions which coincide with

$$z_1, z_2 \text{ in } E, \qquad z_2 - a, z_1 + a \text{ in } E', \qquad z_2 + a, z_1 - a \text{ in } E''.$$

We verify at once that these functions coincide with z_1, z_2 at boundary points of E, since we then have $z_2 - z_1 = \pm a$; it follows easily that they satisfy a Lipschitz condition with a same constant as the functions z_1 and z_2 : for any segment can be decomposed into at most three, each of which has both ends belonging to the closure of the same set E, E', or E'', and in these partial segments the Lipschitz condition for the new functions is identical with the original one for z_1, z_2 or their permutation z_2, z_1.

Finally it is clear that the new functions have the same gradient as the original ones or their permutation, wherever the four gradients exist, and so almost everywhere; and since we have also

$$\tfrac{1}{2}\bar{z}_1 + \tfrac{1}{2}\bar{z}_2 = \tfrac{1}{2}z_1 + \tfrac{1}{2}z_2 \quad \text{and} \quad |\bar{z}_2 - \bar{z}_1| \leqq a,$$

this completes the proof.

3. Integrals over mean surfaces. Comparison of minima

We shall be dealing with surface integrals of functions $f(x, y, z, p, q)$ measurable B but not necessarily continuous. The average $M(g)$ at x, y which constitutes part of the definition of generalized surface, must now be suitably extended to discontinuous functions $g(p, q)$. In the cases that we shall be concerned with, this extension is either trivial, or else sufficiently treated by well-known classical writers.

As in the first note, all surfaces which occur will be supposed Lipschitzian.

In the case of a mean surface S, the average $M(g)$ is of the form $\frac{1}{2}g(p', q') + \frac{1}{2}g(p'', q'')$, where p', q' and p'', q'' denote positions of p, q dependent on x, y, which will be termed the two *possible positions* of p, q at x, y. In accordance with our definition of mean surface, these possible positions are further restricted to be at almost every x, y of the domain A a permutation of the gradients of at least one fixed pair of ordinary surfaces. The integral $F(S)$ of $f(x, y, z, p, q)$ over the mean surface S is obtained by writing

$$\iint_S f(x, y, z, p, q)\, dx\, dy = \iint_A \tfrac{1}{2}\{f(x, y, z, p', q') + f(x, y, z, p'', q'')\}\, dx\, dy,$$

where on the right z denotes the track of S.

The notions of ordinary surface, mean surface, generalized surface give rise to three different problems of minimum, obtained by interpreting correspondingly the class of admissible surfaces with a given boundary for which we seek to make $F(S)$ a minimum. If the corresponding minimal values of $F(S)$ are denoted by m_o, m_m, m_g, it is evident that

$$m_o \geqq m_m \geqq m_g, \tag{3.1}$$

since every ordinary surface is a mean surface (the mean of two coincident, or parallel, ordinary surfaces), while every mean surface is a generalized surface.

The results of the first note[4] show that the inequalities (3.1) reduce to equalities when f is continuous, since the extreme terms then have the same value. On the other hand, it is easily seen that at least one of the inequalities is strict for certain discontinuous functions f: thus if we write $f = -1$ when p, q equals either $-y, x$ or $y, -x$, and $f = 0$ otherwise, we obtain a function $f(x, y, p, q)$ which vanishes almost everywhere on any ordinary surface, but which has the average $M(f) = -1$ at all points of the generalized surface with the track $z = 0$ and the average $M(g)$ at x, y given by the expression $\frac{1}{2}g(-y, x) + \frac{1}{2}g(y, -x)$; so that m_o vanishes while m_g is negative in this case.

We shall show, in the next paragraph, that for a class of functions including among others all those of the form $f(x, y, p, q)$ which are bounded in bounded sets, and in particular the function just considered, the first of the inequalities (3.1) is necessarily an equality, i.e. we have

$$m_o = m_m . \tag{3.2}$$

4. Equality of the minima m_o and m_m

We now come to the first main theorem of this note, which will enable us to treat the ordinary problem on the same footing as the very much simpler generalized problem from the point of view of necessary conditions. This theorem asserts that (3.2) is true under general conditions. Since this equality has already been established for continuous integrands $f(x, y, z, p, q)$, the reader whose interest is limited to classical problems may feel inclined to omit its proof. It should be noted, however, that it is frequently desirable to transform a classical integrand into a discontinuous one by moving the origin of z, p, q to a position which depends on x, y. Such a transformation can be used, for instance, to render certain results of the present note applicable, not only to Lipschitzian, but also to more general surfaces. If we do not insist on these extensions, it is on account of the greater simplicity of the statements for the Lipschitzian case and because we have no evidence that conditions derived from these by means of the above transformation would be essential ones. It is also, perhaps, interesting and instructive to see how the necessary conditions of Euler and Weierstrass, once they have been freed of the formal differential apparatus, can be applied to integrands which are not even continuous.

(4.1) *Suppose firstly that $f(x, y, z, p, q)$ is a function measurable B which is upper-semicontinuous in z when the other variables are fixed;*[5] *and secondly that this function is bounded above*[6] *for any relevant bounded system of x, y, z, p, q. Then, given any mean surface S and real numbers k and ϵ, where $k > F(S)$ and $\epsilon > 0$, there is an ordinary surface S' with the same boundary as S, such that*

---

[4] Young [12], (14.1), p. 102.

[5] This is in particular the case when f is independent of z.

[6] This second condition may be relaxed: it is enough to suppose that in A the function f is less than an integrable function of x, y only. Moreover, when f is independent of z, we require this only in the neighbourhood of the boundary of A.

(4.2) $$F(S') < k,$$

and such that, moreover, its track differs from that of S by at most ϵ, and its Lipschitz constant exceeds that of S by at most ϵ.

Denoting by n an arbitrary positive integer and by E the set of the points x, y of A whose distance from the boundary of A is not less than $1/n$, we can, by (2.1) and McShane's lemma,[7] construct two ordinary surfaces S' and S'' having S for their mean in the set E, in such a manner that the tracks of S' and S'' differ by at most $1/n^2$ in E and coincide with that of S on the boundary of A, while at the same time their Lipschitz constants exceed by at most $1/n$ that of S. We shall suppose that $F(S') \leqq F(S'')$.

Now the tracks z' and z'' of our two surfaces tend to the track z of S as n tends to infinity. Moreover, any point x, y of A becomes a point of E if n is sufficiently large, provided that x, y is not a boundary point; therefore at almost every point of A the gradients p', q' and p'', q'' of our two surfaces are a permutation of the possible positions of p, q for the mean surface S, provided that n is large enough. Hence the average $M(f)$ for S is almost nowhere less than the upper limit as n tends to infinity of the expression

$$\tfrac{1}{2}\{f(x, y, z', p', q') + f(x, y, z'', p'', q'')\},$$

by semicontinuity of f. It follows from a suitable form of Fatou's theorem[8] that

$$\iint_A M(f)\,dx\,dy \geqq \operatorname*{Lim.\,Sup.}_n \iint_A \tfrac{1}{2}\{f(x,y,z',p',q') + f(x,y,z'',p'',q'')\}\,dx\,dy$$

and so, that $k > \frac{1}{2}F(S') + \frac{1}{2}F(S'')$, if n is sufficiently large; clearly this requires $k > F(S')$. The remaining assertions are satisfied if we choose a value of n exceeding ϵ^{-2}. This completes the proof.

5. A convexity property of the minimum

From now on, it will be assumed that the integrand of the problem is a function $f(x, y, p, q)$ which is independent of the variable z. This function will be supposed moreover measurable B and, for the present, bounded above when x, y lies in A and p, q in any bounded set.

Denoting by U a function defined on the boundary of A in such a manner that there exists a Lipschitzian surface whose boundary is given by U, we introduce the functionals $\varphi_o(U)$, $\varphi_g(U)$, $\varphi_o(U, K)$ and $\varphi_g(U, K)$, any one of which we designate shortly $\varphi(U)$. We define these to be absolute minima of our surface-integral for the surfaces with boundary U which belong respectively to the following four classes:

ordinary Lipschitzian surfaces;
generalized Lipschitzian surfaces;
ordinary surfaces with Lipschitz constants less than K;
generalized surfaces with Lipschitz constants less than K.

[7] Young [12], (9.1), p. 93.

[8] Saks [8], (12.10) p. 29, in the form obtained by taking a sequence $a - f_n$ instead of a non-negative sequence f_n.

(5.1) *Each of the minima $\varphi(U)$ is convex in U.*

Before proving this, let us state for completeness, that we term *convex* a functional $Q(U)$ in an arbitrary functional space, if given any two points U' and U'' of this space at which this functional is defined and finite above, and given further any real number θ where $0 \leqq \theta \leqq 1$, the functional is defined at the point $\theta U' + (1 - \theta)U''$ and satisfies

$$Q[\theta U' + (1 - \theta)U''] \leqq \theta Q(U') + (1 - \theta)Q(U''). \tag{5.2}$$

It is known that this definition is equivalent to the following conditions:

the left-hand side of (5.2) is a function of θ bounded above on the unit segment;

the inequality (5.2) holds for $\theta = \frac{1}{2}$.[9]

We shall use this equivalent form of the definition in proving (5.1).

We observe in the first place that if U' and U'' are boundaries of Lipschitzian surfaces belonging to the relevant class, then the same is the case of $\theta U' + (1 - \theta)U''$ when θ lies on the unit segment *and indeed when θ lies on a segment including the unit segment in its interior*. The left-hand side of (5.2), when we choose $Q(U) = \varphi(U)$, is majorized by the values of $F(S)$ for certain surfaces with bounded Lipschitz constants, and so is certainly bounded above as function of θ on the segment in question.

It remains to verify the second condition. Let k be any number exceeding $\frac{1}{2}\varphi(U') + \frac{1}{2}\varphi(U'')$ and therefore exceeding $\frac{1}{2}F(S') + \frac{1}{2}F(S'')$ where S' are certain suitably chosen admissible surfaces with boundaries U' and U'' respectively. The number k then exceeds $F(S)$, where S denotes the mean of S' and S''. But S has the boundary $\frac{1}{2}U' + \frac{1}{2}U''$ and S is either an admissible surface, or can be added to the admissible surfaces without altering our minimum, and so $\varphi(\frac{1}{2}U' + \frac{1}{2}U'') \leqq F(S) < k$. This completes the proof of (5.1).

In the case of the generalized problem, the theorem just proved has a variant which reduces in the *regular* case to a well-known inequality of Haar-Steiner.[10] We denote by $I(z)$ the minimum of $F(S)$ for surfaces S with the track z, and we note that in the regular case $I(z)$ reduces to the value of $F(S)$ for the ordinary surface with this track.

(5.3) *The functional $I(z)$ is convex.*[11]

To prove this, we select an arbitrary real number θ on the unit segment and two Lipschitzian tracks z' and z''. We denote by S' and S'' two admissible generalized surfaces with these tracks, and by $M'(g)$, $M''(g)$ the corresponding averages at x, y; and we write S for the generalized surface with the track $\theta z' + (1 - \theta)z''$ and the average $\theta M' + (1 - \theta)M''$ at x, y. Then

$$I(\theta z' + [1 - \theta]z'') \leqq F(S) = \theta F(S') + [1 - \theta]F(S''),$$

[9] The equivalence of the definitions easily reduces to the case of real functions treated by Hardy, Littlewood and Pólya [6].

[10] Haar [3], p. 227, Inequality (3).

[11] We assert this only for the generalized problem. The same is then evidently true of the ordinary problem *only in the regular case*.

from which it follows, passing to the lower bound for all S' and S'', that

$$I(\theta z' + [1 - \theta]z'') \leqq \theta I(z') + [1 - \theta]I(z'').$$

This completes the proof.

For our purposes, a second variant of (5.1), valid this time for ordinary admissible surfaces as well as for generalized ones, is required. This variant constitutes a minor extension of the original theorem.

6. Prescribed discontinuities[12]

The classical minimum problems for continuous surfaces may be regarded as special cases of corresponding problems for more general surfaces whose discontinuities are specified. We shall require the extension of (5.1) to a problem of this type, in which we admit piecewise Lipschitzian surfaces with prescribed discontinuities along given parallels to the coördinate axes.

By a *piecewise Lipschitzian surface* S (ordinary or generalized) we shall mean the aggregate of a finite number of Lipschitzian portions of surface S_{ik} (ordinary or generalized) defined on partial domains A_{ik} of a subdivision of A by a finite number of parallels to the axes. It will be understood that S is regarded as unaltered by subsequent subdivision of its constituent Lipschitzian portions. By the *surface integral* $F(S)$, we shall mean similarly the sum of the surface integrals $F(S_{ik})$ over these portions, and this definition is again invariant under further subdivision.

By the *discontinuity of* S, we shall mean a function $V(x, y)$ only defined at interior points of A and possibly undefined at a finite number of these, such that V vanishes at interior points of each A_{ik} while on the line separating two adjacent partial domains $A_{i+1,k}$ and A_{ik} or else $A_{i,k+1}$ and A_{ik}, the function V is the difference of the tracks of the corresponding Lipschitzian portions. Besides this discontinuity V, we have a boundary-function U just as in the case of a continuous surface.

It is convenient to combine the two functions U on the boundary of A and V in the interior of A, into a single function W defined throughout A (except possibly at a finite number of points). The function W then specifies the boundary and discontinuity of S.

(6.1) *The functions W which correspond in this way to at least one such piecewise Lipschitzian surface S constitute a vector-space.*

For if S' and S'' are two piecewise Lipschitzian surfaces and W' and W'' specify their boundaries and discontinuities, we can divide A into portions A_{ik} corresponding to continuous portions of both surfaces simultaneously. Denoting by z'_{ik} and z''_{ik} the tracks of these portions we define a piecewise Lipschitzian surface S consisting of the continuous portions of ordinary surfaces

[12] We do not study this problem for its own sake, but in order to obtain machinery for the special case in which there are no discontinuities. In the literature only problems with *unspecified* discontinuities appear to have been treated.

given by the track $az'_{ik} + bz''_{ik}$ on A_{ik}. This surface S then has the boundary and discontinuity specified by the function $W = aW' + bW''$, and this proves (6.1). Moreover, we see that S has in each of its continuous portions a Lipschitz constant not exceeding $|a| + |b|$ times the greatest of those of S' and S'', so that we have the following result:

(6.2) *If W' and W'' correspond to piecewise Lipschitzian surfaces whose continuous portions have Lipschitz constants less than K, then $aW' + bW''$ corresponds to at least one piecewise Lipschitzian surface whose continuous portions have Lipschitz constants less than $(|a| + |b|)K$.*

This being so, we define by analogy with §5 the functionals of W consisting of the absolute minima of $F(S)$ for the piecewise Lipschitzian surfaces S with boundary and discontinuity specified by W which can be subdivided into continuous portions belonging to the same four classes as previously. As these functionals reduce to those of §5 when the variable W reduces to U, i.e. when there is no discontinuity, we shall designate them by the symbols $\varphi_o(W)$, $\varphi_g(W)$, $\varphi_o(W, K)$ and $\varphi_g(W, K)$, and any one by the symbol $\varphi(W)$.

As an immediate consequence of (6.2) we have:

(6.3) *Any two points of the W-space for which $\varphi(W) \neq +\infty$ are inner points of a segment on which $\varphi(W)$ is bounded above.*

Moreover, by exactly the same argument as that used to prove the corresponding inequality in §5, we see that

$$\varphi(\tfrac{1}{2}W' + \tfrac{1}{2}W'') \leqq \tfrac{1}{2}\varphi(W') + \tfrac{1}{2}\varphi(W''). \tag{6.4}$$

Combining (6.3) and (6.4), we deduce:

(6.5) *Each of the minima $\varphi(W)$ is convex in W.*

7. Existence of a "flat support"

The property of convexity of the functionals $\varphi(U)$, $I(z)$, $\varphi(W)$ owes its importance to the fact that it places at our disposal the powerful theorem of Minkowski asserting the existence of a "flat support." We shall follow Banach[13] in proving a form of this theorem valid in general vector spaces which covers our needs.

We consider a space whose elements are abstract vectors X. As is customary, a functional $L(X)$ is termed linear if $L(cX) = cL(X)$ for any constant c, and $L(X + Y) = L(X) + L(Y)$; and since this definition implies that $L(0) = 0$, it is frequently necessary to introduce an additive constant when the analogy with Euclidean space is to be preserved.

The expression $a + L(X)$, where a is a constant and $L(X)$ a linear functional, is termed "*flat support*" at X_0 of a given functional $Q(X)$, if it never exceeds $Q(X)$ and if it takes the value $Q(X_0)$ at X_0.

[13] Banach [1], p. 27, Theorem 1.

One more definition: if G is a vector subspace, we denote by $\{X_0, G\}$ the vector space or subspace which consists of all vectors Y of the form $X + tX_0$ where X lies in G and t is real. It is easy to see that if X_0 does not lie in G, this expression of a vector Y of $\{X_0, G\}$ is *unique*. We shall denote further by $\{X_0, G\}_+$ and $\{X_0, G\}_-$ the systems of vectors Y of the above form with t positive, and with t negative, respectively.

(7.1) *Let $Q(X)$ be a convex functional defined in a vector-space, and let $L(X)$ be a linear functional defined in a vector subspace G and such that throughout G we have $L(X) \leqq Q(X)$; suppose further that X_0 is any point of the vector-space such that $Q(X)$ is defined and finite above for at least one X of each of the sets $\{X_0, G\}_+$ and $\{X_0, G\}_-$. Then we can extend the definition of $L(X)$ in such a manner that this functional be defined, linear, and not greater than $Q(X)$ throughout $\{X_0, G\}$.*

In both the hypothesis and the conclusion of the theorem, the inequality connecting $L(X)$ and $Q(X)$ is regarded as automatically true at points X for which $Q(X)$ is either not defined or not finite above.

In proving (7.1), we may clearly suppose that X_0 does not lie in G. Denoting by k_0 a number to be fixed later, we continue $L(X)$ by writing, for real t and for arbitrary X in G,

$$L(X + tX_0) = L(X) + tk_0,$$

and this continuation is linear in $\{X_0, G\}$, so that we have only to show that it does not exceed $Q(X + tX_0)$. Since this last condition holds for $t = 0$, we need only verify it for $t = a$ and $t = -a'$, where a, a' are positive; it becomes

$$(7.2) \quad \{L(X) - Q(X - a'X_0)\}/a' \leqq k_0 \leqq \{Q(X' + aX_0) - L(X')\}/a$$

where X and X' are arbitrary vectors of G.

It only remains to establish the existence of a finite k_0 which satisfies (7.2) for all a, a', X, X'. Moreover, since Q is defined and finite above *somewhere* in each of the sets $\{X_0, G\}_+$ and $\{X_0, G\}_-$, we see that the upper bound of the expression on the extreme left of (7.2) is not $-\infty$, and the lower bound of the expression on the extreme right is not $+\infty$. The existence of k_0 is thus ensured if we have an inequality between these bounds, which is equivalent to the statement that every value of the extreme left of (7.2) is less than or equal to every value on the extreme right.

Now this last statement, after multiplying by $aa'/(a + a')$, reduces to the inequality

$$\frac{aL(X) + a'L(X')}{a + a'} \leqq \frac{aQ(X - a'X_0) + a'Q(X' + aX_0)}{a + a'}$$

and the latter is satisfied since, if X'' is the vector $(aX + a'X')/(a + a')$ of G (which may also be written $[a(X - a'X_0) + a'(X' + aX_0)]/(a + a')$), the ratio on the left equals $L(X'')$ by linearity, while that on the right is not less than $Q(X'')$ by convexity and so, by hypothesis, not less than $L(X'')$. This completes the proof, which follows closely that given in a special case by Banach.

As a consequence of the theorem just proved, we deduce for our convex functionals $\varphi(W)$, still using only ideas set out by Banach,[14] the following corollary:

(7.3) *Each of the functionals $\varphi(W)$ of §6 has a flat support at any W_0 for which its value $\varphi(W_0)$ is finite.*

To prove this, we write $Q(X)$ for the difference $\varphi(W_0 + X) - \varphi(W_0)$ and, since the latter vanishes for $X = 0$, we need only obtain a linear functional $L(X)$ nowhere exceeding $Q(X)$. We shall construct such an $L(X)$ by transfinite induction, first observing that it clearly exists in the space consisting of the single point 0. Let the points at which Q is defined and finite above be well-ordered, the origin 0 being first. Let X_0 be a subsequent point and let G be the vector subspace consisting of the linear combinations of vectors previous to X_0.

We make the inductive hypothesis that $L(X)$ exists in G with the properties required. By (6.3) the functional Q is bounded above, and so finite above, on a segment containing 0 and X_0 in its interior, and therefore at some point of each of the sets $\{X_0, G\}_+$ and $\{X_0, G\}_-$. By (7.1) we can therefore continue $L(X)$ into $\{X_0, G\}$, and so, by transfinite induction, into the whole of the vector subspace consisting of points which are linear combinations of those at which the functional Q is defined and finite above. This subspace is precisely the whole of our vector space by (6.2).

8. Necessary conditions for an attained minimum

Denoting by U_0 a particular boundary-function, we now propose to examine some consequences of the hypothesis that one of the minima $\varphi(U_0)$ is attained, i.e. that for some surface S_0 having the boundary U_0 and belonging to the appropriate one of the four classes introduced in §5, we have $\varphi(U_0) = F(S_0)$.

We shall modify our original assumption concerning the function $f(x, y, p, q)$, and suppose that it is *not this function but the difference*

$$f(x, y, p, q) - f_0(x, y)$$

which is bounded above when x, y lies in A and p, q lies in a bounded set, where $f_0(x, y)$ denotes the average of f at x, y for the surface S_0 or, in particular, in the case of an ordinary surface, the value of f when we substitute for p, q the gradient of S_0. In addition, we shall suppose, as is really implied in the fact that $F(S_0)$ exists, that $f_0(x, y)$ is integrable in some sense on every portion of A. It is then possible to write $f_0(x, y) = 0$ without effective loss of generality, and we shall do this when convenient in any proof.

We shall denote by $z_0(x, y)$ or simply z_0, the track of S_0, and by W_0 the abstract vector which specifies its boundary U_0 together with the discontinuity zero. We denote further by $z(x, y)$ or simply by z, the track of an arbitrary admissible Lipschitzian surface S, and by $\bar{f}(x, y)$ the average at x, y of the func-

[14] Banach [1], p. 29, Corollary.

tion $f(x, y, p, q)$ for this surface. Finally we denote by Δ an arbitrary interval of the x, y plane.

(8.1) *There exists a linear functional $L(z; \Delta)$ of z, additive in Δ dependent only on the values of z on the boundary of $A \cdot \Delta$, such that*

$$\text{(8.2)} \qquad \iint_{A\cdot\Delta} f(x, y)\, dx\, dy \geqq \iint_{A\cdot\Delta} f_0(x, y)\, dx\, dy + L(z; \Delta) - L(z_0; \Delta)$$

for every admissible surface, (i.e. for every Lipschitzian surface of the appropriate class), whatever its boundary U may be.

(8.3) *In particular, taking for Δ an interval including A, there is a linear functional $L(z)$ depending only on the boundary-values U of the track z of the surface S, such that*

$$\text{(8.4)} \qquad F(S) \geqq F(S_0) + L(z) - L(z_0).$$

It will be observed that this last statement includes as a special case the minimizing hypothesis from which we started, namely that

$$F(S) \leqq F(S_0) \quad \text{when} \quad U = U_0\,,$$

since we then have $L(z) = L(z_0)$. So that (8.3), and *a fortiori* (8.1), are theorems in which the hypotheses are exploited to the full.

We may suppose $f_0(x, y) = 0$, so that $f(x, y, p, q)$ now satisfies the hypotheses of §5. We have then $\varphi(W_0) = 0$, so that by (7.3) there is a linear functional of $\overline{W} - W_0$ nowhere exceeding $\varphi(\overline{W})$, and *a fortiori* not exceeding the values of $F(\bar{S})$ for a discontinuous $\bar{S}$ of the appropriate type whose discontinuity and boundary are specified by $\overline{W}$. We select for $\bar{S}$ a surface which coincides in Δ with an admissible continuous surface S and outside Δ with the minimizing surface S_0. The linear functional of $\overline{W} - W_0$ obtained becomes a functional of the track of S and of the interval Δ, clearly dependent only on the boundary-values of this track z in $A \cdot \Delta$, and which we denote by $L(z - z_0, \Delta)$. The functional $L(z, \Delta)$ is moreover linear in z and additive in Δ, by construction; furthermore it satisfies (8.2) since the left-hand side is $F(\bar{S})$. This completes the proof.

We shall now proceed to derive from (8.2) further information regarding the functional $L(z, \Delta)$ whose existence is asserted there.

(8.5) *There exist bounded measurable functions of x, y, the functions $P(x, y)$ and $Q(x, y)$ say, such that*

$$\text{(8.6)} \qquad L(z, \Delta) = \iint_{A\cdot\Delta} (P \cdot z_y - Q \cdot z_x)\, dx\, dy.$$

In proving this, we again suppose $f_0(x, y) = 0$. Moreover we may restrict ourselves, by homogeneity, to tracks z of surfaces whose Lipschitz constants are sufficiently small for $z + z_0$ to be the track of an admissible surface. The function $f(x, y, p, q)$ being now bounded above for the relevant surfaces, it

follows by applying (8.2) to surfaces with the track $z + z_0$, that $L(z, \Delta)$ cannot exceed $N \cdot |\Delta|$, where N is a constant and $|\Delta|$ is the area of Δ. We can therefore express $L(z, \Delta)$ as the integral of its derivative in Δ, the latter existing almost everywhere for fixed z.

In particular, when z is a linear function $ax + by + c$, the functional L, which is clearly unaltered by passing to a parallel surface, will be of the form $aA + bB$ where A and B are functions of Δ possessing bounded derivatives in Δ almost everywhere. Denoting by $-Q$ and $+P$ these derivatives at x, y, and writing for instance $P = Q = 0$ when they do not exist, the formula to be proved clearly holds for linear tracks z.

In order to establish the formula generally, we denote its right-hand side by $L_1(z, \Delta)$. We denote further by z_1 a track with the same boundary as z on $A \cdot \Delta$ and with a Lipschitz constant increased by as little as we please, whose gradient in a subset of Δ of measure as close as we please to Δ differs arbitrarily little from that of z and is constant in a neighborhood of each point of this subset.[15] Now the difference of $L_1(z_1, \Delta)$ and $L_1(z, \Delta)$ is arbitrarily small since it is less than a constant multiple of the integral of the absolute difference of the gradients of z and z_1. Moreover $L(z_1, \Delta)$ and $L_1(z_1, \Delta)$ have the same derivative outside small measure, and their derivatives elsewhere are bounded; therefore their difference also is as small as we please. Finally $L(z, \Delta)$ and $L(z_1, \Delta)$ coincide since z and z_1 have the same boundary on $A \cdot \Delta$. Combining these results, we see that $L(z, \Delta)$ and $L_1(z, \Delta)$ differ arbitrarily little, and so coincide. This completes the proof of (8.5).

These results can now be sharpened still further by applying the following lemma due to Haar[16] and Schauder:[17]

(8.7) *In order that the integral* $\iint_A (P \cdot z_y - Q \cdot z_x)\, dx\, dy$, *where* P, Q *are bounded measurable functions of* x, y, *shall take the value* 0 *whenever* z *is a Lipschitzian function vanishing on the boundary of* A, *it is both necessary and sufficient that there exist a Lipschitzian function* Z *whose gradient is almost everywhere* P, Q.

It is assumed in the above lemma that A is a bounded simply-connected domain.[18]

As a consequence of these results, our main theorem becomes:

(8.8) *In order that* $F(S_0) = \varphi(U_0)$, *where* S_0 *satisfies the remaining hypotheses of the beginning of this paragraph, it is necessary and sufficient that there exist a Lipschitzian function* Z *such that the following inequality be verified for every admissible surface* (*whose track* z *has an arbitrary boundary*):

$$(8.9) \qquad \iint_{A \cdot \Delta} \{\bar{f}(x, y) - f_0(x, y) - Z_x(z_y - z_{0_y}) + Z_y(z_x - z_{0_x})\}\, dx\, dy \geqq 0.$$

[15] The existence of such a track is ensured by (12.1) p. 98 of the first note, Young [12].

[16] Haar [2], [4].

[17] Schauder [9].

[18] Our results have been mainly stated for convex domains but hold generally. At this stage however they extend only to simply-connected domains.

9. Localized form of the conditions. Connection with Weierstrass' condition and with the Euler-Haar equations

Following Bonnet and Haar, we term *adjoint*[19] of the minimizing surface S_0, the ordinary surface with the track $Z(x, y)$ whose existence is established in (8.8). Our object is to reduce (8.9) to the inequality

$$f(x, y, p, q) - f_0(x, y) - P\cdot(q - q_0) + Q\cdot(p - p_0) \geqq 0, \tag{9.1}$$

where p_0, q_0 and P, Q are the gradients at x, y of the tracks z_0 and Z. The validity of (9.1) for all p, q will be termed *the condition* (W, E) *at* x, y; similarly, its validity for all p, q interior to the circle of radius K and center at the origin, will be termed *the condition* $(W, E)_K$. These terms are similar to those adopted by the writer in the case of curves[20] and the most elementary considerations show that for a function $f(x, y, p, q)$ possessing partial derivatives in p and q at x, y (and in particular for the integrands of all classical problems), the condition (W, E) is wholly equivalent to the simultaneous validity of

$$f_{p_0} = -Q, \qquad f_{q_0} = P,$$

$$f(x, y, p, q) - f_0(x, y) - f_{p_0}\cdot(p - p_0) - f_{q_0}\cdot(q - q_0) \geqq 0.$$

The first two relations are Haar's form of the Euler equations, while the third is Weierstrass' condition.

A set E of values of x, y, p, q will be termed *superficially negligible*, if given any ordinary Lipschitzian surface S with the track $z(x, y)$, the set of the points x, y, p, q of E at which p, q is the gradient of this track has a projection in the x, y plane of measure zero. In view of this definition, it is clear that an alteration of $f(x, y, p, q)$ in such a set of values of x, y, p, q cannot affect the value of $F(S)$ for ordinary Lipschitzian surfaces. On the other hand, the example considered in §3 shows that such an alteration may affect the values of $F(S)$ for generalized S. This difference between the ordinary and the generalized problems will cause slight differences between the corresponding localized forms of the conditions for an attained minimum, which we state in (9.2) and (9.3) below.

We begin with the ordinary problem.

(9.2) *With the hypotheses of* (8.8), *where we now suppose* S_0 *an ordinary surface, a necessary and sufficient condition for an ordinary minimum is that the inequality* (9.1) *be verified for a corresponding adjoint surface with the gradient* P, Q *except at most in a superficially negligible set of* x, y, p, q.

This is clearly equivalent to the condition that the integrand of left-hand side of (8.9) be non-negative for every admissible surface except at most in a set of x, y of plane measure 0; and this in its turn is clearly equivalent to (8.9) itself. A similar result holds for the ordinary minimum in the class of *surfaces with Lipschitz constants less than* K.

---

[19] Haar [5].

[20] Young [11].

The corresponding results for the generalized problem are more complete and also not so easy to deduce from (8.8).

(9.3) *With the hypotheses of* (8.8), *a necessary and sufficient condition for a minimum among generalized surfaces whose Lipschitz constants are either unrestricted or less than a fixed* K, *is that at almost points* x, y *of* A *the minimizing surface* S_0 *satisfy either the condition* (W, E), *or else the condition* $(W, E)_K$.

The main difference between this statement and the one concerning ordinary minima is that more information is given here about the exceptional set of x, y, p, q for which (9.1) need not hold, *its projection in the* x, y *plane being this time a null set.* Incidentally the combination of the two results will show that if an ordinary minimum is attained, then by altering f for at most a superficially negligible set of x, y, p, q we ensure that the same minimizing surface provides a generalized minimum.

In proving (9.3), we may suppose as usual $f_0(x, y) = 0$. We shall denote by K_0 any number exceeding the Lipschitz constant of S_0 and less than the upper bound of the Lipschitz constants of all admissible surfaces. We shall denote further by $f^*(x, y, p, q)$ the lower bound of the expression

$$\sum_{k=1}^{N} a_k f(x, y, p_k, q_k) \tag{9.4}$$

as function of the integer N and of the a_k, p_k, q_k where $k = 1, \cdots, N$ when these variables are restricted by the conditions

$$a_k \geqq 0, \qquad p_k^2 + q_k^2 \leqq K_0^2,$$

$$\sum_{k=1}^{N} a_k = 1, \qquad \sum_{k=1}^{N} a_k p_k = p, \qquad \sum_{k=1}^{N} a_k q_k = q. \tag{9.5}$$

It is easy to see that the lower bound of (9.4) is unaltered *if we impose the additional restriction* $N \leqq 3$. For if $N > 4$, there exists a system of numbers b_k, such that not all b_k vanish, which fulfil the equations

$$\sum_{k=1}^{N} b_k = \sum_{k=1}^{N} b_k p_k = \sum_{k=1}^{N} b_k q_k = 0;$$

we can arrange, by changing the signs if necessary throughout, that

$$\sum_{k=1}^{N} b_k f(x, y, p_k, q_k) \leqq 0,$$

and, by multiplying through by a suitable positive constant (remembering that at least one b_k must be negative), that the *least* of the numbers

$$a_k' = a_k + b_k$$

is equal to zero. It follows that we can reduce the value of N without increasing (9.4), by taking as new system of a_k, p_k, q_k those of the a_k', p_k, q_k for which a_k' does not vanish, and so, by induction we can make $N \leqq 3$.

We shall require, for the purpose of proving (9.3), the following properties of the function $f^*(x, y, p, q)$:

(9.6) (i) *For constant x, y the function $f^*(x, y, p, q)$ is convex in p, q.*
(ii) *For constant p, q the function $f^*(x, y, p, q)$ is measurable in x, y.*

The first of these is an immediate consequence of the definition. To prove (ii), it is sufficient to prove that the set of x, y at which any given number c is not greater than f^* is measurable; this set is clearly the projection in the x, y plane of a set of x, y, p_k, q_k, a_k $(k = 1, 2, 3)$ for which (9.5) holds and for which the expression (9.4) is greater than or equal to c; as the projection of a set measurable B, it is therefore itself[21] measurable (though not in general measurable B). This completes the proof of (9.6) (ii).

We shall still require one further lemma:

(9.7) *Given any finite measurable function $h(x, y)$ greater than $f^*(x, y, p, q)$, there exists a generalized surface with constant gradient p, q and with a Lipschitz constant not exceeding K_0, such that $M(f) < h$ for almost all x, y of A.*

To prove this, we consider the Borel set of x, y, p_k, q_k, a_k $(k = 1, 2, 3)$ in space of eleven dimensions, for which (9.5) holds and for which the expression (9.4) is less than $h(x, y)$. The projection of this set in the x, y plane contains every point of A, therefore[22] the set itself contains the graph of a system of nine functions p_k, q_k, a_k $(k = 1, 2, 3)$ of x, y measurable in A. Substituting these functions in (9.4), this expression becomes an average $M(f)$ whose value is a measurable function of x, y less than $h(x, y)$.

We are now in a position to establish (9.3). For any p, q of magnitude not exceeding K_0, it is clear that f^* is a finite function of x, y not exceeding f, provided that we exclude if necessary a null set of x, y where $f = -\infty$ (if this last set were not a null set, it would still have positive measure when p, q is replaced by the gradient of S_0). Moreover, by excluding a suitable null set, we may suppose f^*, for the given constant p, q, to be measurable B in x, y. Choosing $h(x, y)$ to be $f^*(x, y, p, q) + \epsilon$ in this set, and say $f(x, y, p, q) + \epsilon$ in the complementary null set, it clearly follows from (9.7) and (8.9) that for almost all x, y of A,

$$f^* + \epsilon \geqq f_0 + Z_x(q - z_{0_y}) - Z_y(p - z_{0_x}).$$

From this inequality, making ϵ describe a sequence with limit 0 and making p, q describe all rational points of magnitude not exceeding K_0, we deduce that outside a fixed null set, the inequality

$$f^* \geqq f_0 + Z_x(q - z_{0_y}) - Z_y(p - z_{0_x})$$

holds for all such rational p, q. This same inequality, and *a fortiori* the inequality (9.1), then holds for the same set of x, y and *all* p, q whose magnitude

[21] Lusin [7], p. 144 & p. 152; cf. also Saks [8], p. 47–50.
[22] By the argument of Young [11], p. 237, lemma (5.2) & p. 256, 257.

does not exceed K_0, since f^* is continuous in p, q on account of its finiteness and convexity. Finally, making K_0 describe a sequence with the limit ∞ or K, we see that (9.1) must still hold outside a null set of x, y for all relevant p, q, and this is what we asserted. The converse being a trivial consequence of (8.8), this completes the proof.

Let us remark in conclusion that the converse part of these theorems can be strengthened: A surface S_0 satisfying the condition (W, E) with a Lipschitzian adjoint, provides a minimum in a wider class of surfaces than those originally admitted, namely for all surfaces whose track z in such that the integral $\iint_A (Z_x z_y - Z_y z_x)\, dx\, dy$ depends only on the boundary values of the variable track z.

It has been proved in the case of a rectangle A, and the proof extends easily enough to the case in which A is a convex region, that the integral in question, where Z is a given Lipschitzian function, has this property for any z absolutely continuous in the sense of Tonelli.[23]

CAPETOWN, SOUTH AFRICA.

REFERENCES

[1] BANACH, S. *Théorie des opérations linéaires*, Monografie Matematyczne, I (Warszawa 1932).

[2] HAAR, A. *Über die Variation der Doppelintegrale*, Journal für die reine und angewandte Mathematik, 146 (1919), pp. 1–18.

[3] HAAR, A., *Über reguläre Variationsprobleme*, Acta Litterarum ac Scientiarum (Szeged), 3 (1927), pp. 224–234.

[4] HAAR, A. *Über das Plateau'sche Problem*, Mathematische Annalen, 97 (1927), pp. 124–158.

[5] HAAR, A. *Über adjungierte Variationsprobleme und adjungierte Extremalflächen*, Mathematische Annalen, 100 (1928), pp. 481–502.

[6] HARDY, G. H., LITTLEWOOD, J. E., AND PÓLYA, G. *Inequalities*, (Cambridge 1934).

[7] LUSIN, N. *Leçons sur les Ensembles Analytiques et leurs applications*, Collection de monographies Borel (Paris 1930).

[8] SAKS, S. *Theory of the integral*, Monografie Matematyczne, VII (Warszawa 1937).

[9] SCHAUDER, J. *Über die Umkehrung eines Satzes aus der Variationsrechnung*, Acta Litterarum ac Scientiarum (Szeged), 4 (1928), pp. 28–50.

[10] STEINER, J. *Gesammelte Werke*, Band II.

[11] YOUNG, L. C. *Necessary conditions in the calculus of variations*, Acta Mathematica, 69 (1938) p. 239–258.

[12] YOUNG, L. C. *Generalized surfaces in the calculus of variations, I. Generalized Lipschitzian surfaces.* Annals of Mathematics, 43 (1942), pp. 84–103.

[13] YOUNG, W. H. *On a formula for an area*, Proceedings of the London Mathematical Society, (2) 18 (1919), pp. 339–374.

[14] YOUNG, W. H. *On a new set of conditions for a formula for an area*, Proceedings of the London Mathematical Society, (2) 21 (1923), pp. 75–94.

[23] Young, W. H. [13], [14]. The case in which the derivative in one variable of one of the functions Z, z and the derivative in the other variable of the other function have associated summabilities is also treated.

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

THE GEOMETRY OF ISOTROPIC SURFACES

BY SHIING-SHEN CHERN

(Received May 23, 1941; revised October 10, 1941)

Introduction

The geometry in a space in which there is given a family of varieties has been the object of extensive researches. The geometry of paths[1] corresponds to the case of a $(2n-2)$-parameter family of curves in an n-dimensional space, defined by a system of differential equations of the second order of a certain type. It may be regarded as a natural generalization of projective geometry, since Cartan has proved that of all the projective connections having the same geodesics there exists one, and only one, which is characterized by intrinsic properties and which he called normal.[2] Recently, M. Hachtroudi[3] proved that in a three-dimensional space with a three-parameter family of surfaces there can be defined, in an intrinsic way, one and only one projective connection with the contact elements as the elements of the space. It is the aim of the present paper to study the geometry of a space in which there is given a two-parameter family of surfaces. Our main results may be summarized in the following two theorems:

THEOREM 1. *Suppose there be given in a three-dimensional space a two-parameter family of surfaces $\{\Sigma\}$ such that the tangent planes at a point of the surfaces of the family through the point do not pass through a fixed direction. Then there can be defined in the space one and only one four-dimensional Weyl geometry, which has the contact elements of the surfaces as its "points" and possesses certain intrinsic properties.*

THEOREM 2. *There exists a point transformation carrying one family of surfaces $\{\Sigma\}$ into another $\{\bar{\Sigma}\}$ when and only when the two four-dimensional Weyl geometries are equivalent.*

The paper is divided into four sections. In §1 we give the definition of a Weyl space of elements[4] and of some of its fundamental notions. The determination of the Weyl space from the family of surfaces is then given in §2. In §3 we show that the so-defined Weyl geometry naturally leads to a solution of the problem of equivalence, i.e., the problem of deciding when two such families are equivalent under point transformations. The important particular

---

[1] Cf., for instance, O. Veblen and T. Y. Thomas, *The geometry of paths*, Trans. Amer. Math. Soc., vol. 25 (1923), pp. 551–608.

[2] E. Cartan, *Sur les variétés à connexion projective*, Bull. de la Soc. Math. de France, t. 52 (1924), pp. 205–241.

[3] M. Hachtroudi, *Les espaces d'éléments à connexion projective normale*, Actualités Scientifiques et Industrielles, no. 565, Paris, 1937.

[4] Cf. E. Cartan, *La méthode du repère mobile, la théorie des groupes continus, et les espaces généralisés*, Actualités Scientifiques et Industrielles, Paris, 1935.

case by which the Weyl space reduces to a three-dimensional point space is discussed in §4.

The extension of these results to a space of n-dimensions will be treated in a subsequent paper.

1. Weyl Space of Elements

Consider a three-dimensional space S with the coordinates x, y, z. To each point M of the space we attach three vectors I_1, I_2, I_3, such that their scalar products satisfy the conditions

$$(1) \qquad \begin{cases} I_1 I_3 = -I_2^2 \neq 0, \\ I_1^2 = I_3^2 = I_1 I_2 = I_2 I_3 = 0. \end{cases}$$

The totality of the point M and the vectors I_1, I_2, I_3 through M satisfying (1) will be called a *frame*. We say that the space is a Weyl space or that a Weyl connection is defined in the space, if a law of infinitesimal displacement of the following form is given:

$$\begin{cases} dM = \omega_1 I_1 + \omega_2 I_2 + \omega_3 I_3, \\ dI_1 = \omega_{11} I_1 + \omega_{12} I_2 + \omega_{13} I_3, \\ dI_2 = \omega_{21} I_1 + \omega_{22} I_2 + \omega_{23} I_3, \\ dI_3 = \omega_{31} I_1 + \omega_{32} I_2 + \omega_{33} I_3, \end{cases}$$

where the ω's are Pfaffian forms in x, y, z. On differentiating (1) and making use of the above equations, we find

$$\omega_{13} = 0, \qquad \omega_{31} = 0, \qquad \omega_{11} + \omega_{33} - 2\omega_{22} = 0,$$
$$\omega_{12} - \omega_{23} = 0, \qquad \omega_{21} - \omega_{32} = 0.$$

Hence it turns out to be more convenient to put

$$\tau = \omega_{12}, \qquad \pi_1 = \omega_{11}, \qquad \pi_2 = \omega_{22}, \qquad \pi_3 = \omega_{21},$$

and to write the equations of infinitestimal displacement in the form:

$$(2) \qquad \begin{cases} dM = \omega_1 I_1 + \omega_2 I_2 + \omega_3 I_3, \\ dI_1 = \pi_1 I_1 + \tau I_2, \\ dI_2 = \pi_3 I_1 + \pi_2 I_2 + \tau I_3, \\ dI_3 = \pi_3 I_2 + (-\pi_1 + 2\pi_2) I_3. \end{cases}$$

To serve our later purpose we shall generalize the notion of a Weyl space to that of a *Weyl space of elements*, by which we shall mean the following: We attach to each point of the space a one-parameter family of contact elements with

the point as origin. A contact element is then defined by four coordinates x, y, z, t, where x, y, z are the coordinates of the point and t fixes the plane. By attaching a frame to each contact element and supposing the ω_1, ω_2, ω_3, $\tau, \pi_1, \pi_2, \pi_3$ in (2) to be Pfaffian forms in x, y, z, t, we shall get a Weyl space of elements. To define a Weyl space of elements it is thus necessary to give the one-parameter family of contact elements through each point and the seven Pfaffian forms in (2).

It will be advantageous to interpret the four-parameter family of contact elements to be the "points" of a four-dimensional auxiliary space S'. Given a curve in S' defined by the equations

$$x = x(\lambda), \qquad y = y(\lambda), \qquad z = z(\lambda), \qquad t = t(\lambda), \tag{3}$$

we can express the Pfaffian forms in (2) as Pfaffian forms in λ. Then equations (2) become a system of ordinary differential equations. By a well-known theorem in the theory of differential equations, we see that when the initial frame $M^{(0)}I_1^{(0)}I_2^{(0)}I_3^{(0)}$ corresponding to an initial value $\lambda = \lambda_0$ is given, the family of frames $MI_1I_2I_3$ satisfying (2) is uniquely determined (with respect to $M^{(0)}I_1^{(0)}I_2^{(0)}I_3^{(0)}$). In this case, we say that *the family of frames is developed, along the curve* (3), *in the Euclidean space defined by the frame* $M^{(0)}I_1^{(0)}I_2^{(0)}I_3^{(0)}$. If we join a point P_0 of S' to a point P of S' by two different curves and develop, in the Euclidean space of the frame attached to P_0, the two families of frames along these two curves, the frames obtained corresponding to the frame at P will in general be different.

Instead of studying the development of the frame at P on the frame at P_0 along two different curves, we shall consider an infinitesimal parallelogram with P_0 as a vertex and with two directions d and δ through P_0 as its two sides. Such a parallelogram is called a *cycle*. If we develop the family of frames first along the side d and then along δ or vice versa, the frames obtained will be different, so that there exists an infinitesimal transformation carrying one frame to the other. This infinitesimal transformation is called the *infinitesimal transformation associated to the cycle*. By a classical method,[5] it is easy to show that the infinitesimal transformation associated to a cycle formed by the directions d and δ is given by equations of the form

$$\begin{cases} \nabla M = d\delta M - \delta dM = \Omega_1 I_1 + \Omega_2 I_2 + \Omega_3 I_3, \\ \nabla I_1 = d\delta I_1 - \delta dI_1 = \Pi_1 I_1 + \mathrm{T} I_2, \\ \nabla I_2 = d\delta I_2 - \delta dI_2 = \Pi_3 I_1 + \Pi_2 I_2 + \mathrm{T} I_3, \\ \nabla I_3 = d\delta I_3 - \delta dI_3 = \Pi_3 I_2 + (-\Pi_1 + 2\Pi_2) I_3, \end{cases} \tag{4}$$

[5] Cf., for example, E. Cartan, *Leçons sur la géométrie des espaces de Riemann*, Paris, 1928, Chap. VII.

where, with the notations of Cartan's exterior calculus,[6] we have

$$(5)\qquad \begin{cases} \Omega_1 = \omega_1' + [\pi_1\omega_1] + [\pi_3\omega_2], \\ \Omega_2 = \omega_2' + [\tau\omega_1] + [\pi_2\omega_2] + [\pi_3\omega_3], \\ \Omega_3 = \omega_3' + [\tau\omega_2] - [\pi_1\omega_3] + 2[\pi_2\omega_3], \\ \mathrm{T} = \tau' - [\pi_1\tau] + [\pi_2\tau], \\ \Pi_1 = \pi_1' + [\pi_3\tau], \\ \Pi_2 = \pi_2', \\ \Pi_3 = \pi_3' + [\pi_1\pi_3] - [\pi_2\pi_3], \end{cases}$$

so that T, Ω_i, Π_i $(i = 1, 2, 3)$ are exterior quadratic differential forms in x, y, z, t.

Suppose the frame attached to each contact element be so chosen that M coincides with the point of the element and I_1, I_2 belong to the plane of the element. Then the contact element will be fixed, if and only if

$$\omega_1 = \omega_2 = \omega_3 = \tau = 0.$$

It follows that ω_1, ω_2, ω_3, τ are linear combinations of dx, dy, dz, dt and are themselves linearly independent. We can therefore put

$$(6)\qquad \begin{cases} \Omega_i = P_{i1}[\omega_2\omega_3] + P_{i2}[\omega_3\omega_1] + P_{i3}[\omega_1\omega_2] + \sum_{k=1}^{3} Q_{ik}[\omega_k\tau], & i = 1, 2, 3, \\ \Pi_i = R_{i1}[\omega_2\omega_3] + R_{i2}[\omega_3\omega_1] + R_{i3}[\omega_1\omega_2] + \sum_{k=1}^{3} S_{ik}[\omega_k\tau], & i = 1, 2, 3, \\ \mathrm{T} = P_1[\omega_2\omega_3] + P_2[\omega_3\omega_1] + P_3[\omega_1\omega_2] + \sum_{k=1}^{3} Q_k[\omega_k\tau]. \end{cases}$$

The 42 coefficients in (6) are the components of the "tensor of curvature and torsion" of the space.

The Weyl space of elements as defined above may be studied from two different points of view. We may either regard the space as a genuine four-dimensional space (x, y, z, t) and study its properties invariant under the group (or, more precisely, the pseudo-group) of transformations

$$(7)\qquad \bar{x} = \bar{x}(x, y, z, t), \quad \bar{y} = \bar{y}(x, y, z, t), \quad \bar{z} = \bar{z}(x, y, z, t), \quad \bar{t} = \bar{t}(x, y, z, t),$$

or we may study the properties invariant under the group of transformations

$$(8)\qquad \bar{x} = \bar{x}(x, y, z), \quad \bar{y} = \bar{y}(x, y, z), \quad \bar{z} = \bar{z}(x, y, z), \quad \bar{t} = \bar{t}(x, y, z, t).$$

[6] An introduction to the notions of exterior multiplication and exterior differentiation is given in: E. Cartan, *Leçons sur les invariants intégraux*, Paris, 1922.

For our purpose we shall restrict ourselves to the latter point of view and shall call a property *intrinsic*, if it is invariant under transformations of the form (8). In particular, it follows that the concept of a point in S is an intrinsic property.

2. The Two-Parameter Family of Surfaces and Its Intrinsic Weyl Geometry

Suppose there be given in the space S a family $\{\Sigma\}$ of surfaces depending on two parameters. We shall study the properties of $\{\Sigma\}$, which are not affected by an arbitrary transformation of coordinates

$$\bar{x} = \bar{x}(x, y, z), \qquad \bar{y} = \bar{y}(x, y, z), \qquad \bar{z} = \bar{z}(x, y, z), \tag{9}$$

To make this study it will be convenient to introduce the four-dimensional auxiliary space S' formed by the contact elements of the surfaces of the family $\{\Sigma\}$. By the use of a parameter t the contact elements of $\{\Sigma\}$ having the same origin (x, y, z) can be defined by an equation of the form

$$\theta \equiv A(x, y, z, t)dx + B(x, y, z, t)dy + C(x, y, z, t)dz = 0, \tag{10}$$

where A, B, C are not all zero. The parameter t serves to distinguish the contact elements through the point and is arbtrary to the degree that it may be submitted to a transformation of the form

$$\bar{t} = \bar{t}(x, y, z, t). \tag{11}$$

In the auxiliary space S' of our contact elements with the coordinates x, y, z, t, each surface Σ of the family is represented by a two-dimensional variety. All these two-dimensional variaties fill up the space S' in the sense that through every point (in a certain neighborhood) of S' there passes one, and only one, such variety. The family $\{\Sigma\}$ of surfaces can therefore be defined by a completely integrable Pfaffian system of the form

$$\theta = 0, \qquad \theta_1 \equiv A_1dx + B_1dy + C_1dz + Edt = 0, \qquad E \neq 0, \tag{12}$$

in the four variables x, y, z, t. The complete integrability of the system is expressed by the fact that the exterior derivatives θ', θ_1' are congruent to zero, mod. θ, θ_1.

By this representation of the family $\{\Sigma\}$ of surfaces in S as a family of surfaces in S' the properties of the original family unaffected by transformations of the form (9) are represented by properties of the corresponding family in S', which remain unaltered under the transformations (9), (11), and conversely. But these are exactly the intrinsic properties in S' in the sense defined in §1.

Suppose, without loss of generality, that $C \neq 0$, so that we may simplify the system (12) to the form

$$\begin{cases} \theta \equiv dz - pdx - qdy = 0, \\ \theta_1 \equiv dt - rdx - sdy = 0, \end{cases} \tag{13}$$

where p, q, r, s are functions of x, y, z, t. We introduce the notation

$$(14)\qquad \begin{cases} \dfrac{d}{dx} = \dfrac{\partial}{\partial x} + p\dfrac{\partial}{\partial z} + r\dfrac{\partial}{\partial t}, \\ \dfrac{d}{dy} = \dfrac{\partial}{\partial y} + q\dfrac{\partial}{\partial z} + s\dfrac{\partial}{\partial t}. \end{cases}$$

With this notation we have, for any function F in the variables x, y, z, t, the formula

$$(15)\qquad dF = \frac{dF}{dx}dx + \frac{dF}{dy}dy + \frac{\partial F}{\partial z}\theta + \frac{\partial F}{\partial t}\theta_1,$$

and the integrability conditions of the system (13) can be written in the following simple form:

$$(16)\qquad \frac{dp}{dy} - \frac{dq}{dx} = 0, \qquad \frac{dr}{dy} - \frac{ds}{dx} = 0.$$

We now proceed to prove that when a two-parameter family of surfaces (13) is given in S such that

$$(17)\qquad P \equiv p_t q_u - q_t p_u \neq 0,$$

there can be defined, in an intrinsic way, a Weyl geometry of elements in S. As it is easy to see, *the vanishing of P signifies that the contact elements through a point of S pass through a fixed direction.*

To each contact element formed by the point (x, y, z) and the plane $\theta = 0$ we attach a frame $MI_1I_2I_3$ *adapted to it* by supposing the following conditions to be satisfied:

a) M coincides with the point (x, y, z).

b) I_1 is along the line of intersection of the plane and a neighboring plane through M, i.e., the line defined by the equations

$$(18)\qquad \theta = 0, \qquad p_t dx + q_t dy = 0.$$

We shall call this line the *characteristic* of the element.

c) I_2 lies on the plane $\theta = 0$.

If $MI_1I_2I_3$ is such a frame, the most general frame $M^*I_1^*I_2^*I_3^*$ satisfying the same conditions is given by

$$(19)\qquad \begin{cases} M^* = M, \qquad I_1 = uI_1^*, \qquad I_2 = wI_2^* + \dfrac{uw}{w}I_1^*, \\ I_3 = \dfrac{w^2}{u}I_3^* + vI_2^* + \dfrac{1}{2}\dfrac{uv^2}{w^2}I_1^*, \end{cases}$$

where u, v, w are parameters, with $uw \neq 0$.

We now suppose our Weyl geometry to have the following two intrinsic properties:

1) Points are developed into points, i.e., the contact elements having the same origin are developed into the contact elements having the same origin.

2) The contact elements belonging to a fixed surface of the family are developed into contact elements having the same plane.

Let us express the conditions for these two properties analytically and see how far the Weyl geometry is thus defined. When the point M is fixed in space, we have

$$dx = dy = dz = 0.$$

Hence condition 1 is expressed by the relation

$$dM = 0 \pmod{dx, dy, dz},$$

i.e., ω_1, ω_2, ω_3 are linear combinations of dx, dy, θ. Similarly, condition 2 is expressed by the fact that ω_3, τ are linear combinations of θ and θ_1. It follows in particular that ω_3 is a multiple of θ. On the other hand, the condition b for the choice of the frame means that ω_2 is a linear combination of θ and $p_t dx + q_t dy$.

With a definite system of coordinates x, y, z, t we may make use of the parameters u, v, w in (19) to simplify the expressions for ω_1, ω_2, ω_3. It is easy to verify that we may choose the frame attached to each contact element such that we have

$$(20) \qquad \begin{cases} \omega_2 = -(p_t dx + q_t dy) + \eta\theta, \\ \omega_3 = \theta, \end{cases}$$

where η remains undetermined. The most general change of frame leaving the conditions a, b, c and the form of the equations (20) unaltered is given by

$$(21) \qquad I_1 = I_1^*, \qquad I_2 = I_2^* + vI_1^*, \qquad I_3 = I_3^* + vI_2^* + \tfrac{1}{2}v^2 I_1^*,$$

v being a parameter.

In order to impose further conditions on the Weyl space, we are led to consider two particular kinds of cycles. The first kind is one whose directions d and δ displace the point of the contact element in the same direction and is analytically characterized by the conditions

$$\frac{\omega_1(d)}{\omega_1(\delta)} = \frac{\omega_2(d)}{\omega_2(\delta)} = \frac{\omega_3(d)}{\omega_3(\delta)},$$

or by

$$(22) \qquad [\omega_1\omega_2] = [\omega_2\omega_3] = [\omega_3\omega_1] = 0.$$

The second may be described as one displacing the surface of the contact element in the same direction and will be defined by

$$\frac{\omega_3(d)}{\omega_3(\delta)} = \frac{\tau(d)}{\tau(\delta)},$$

or by

$$(23) \qquad [\omega_3\tau] = 0.$$

The property of a cycle to be either one of the two kinds is intrinsic.

We now suppose our Weyl space to possess the following three further properties:

3) The infinitesimal transformation associated to every cycle displaces the point M in the direction of I_1.

4) The infinitesimal transformation associated to every cycle whose directions d and δ displace the surface of the contact element in the same direction leaves the point M invariant.

5) The infinitesimal transformation associated to every cycle whose directions d and δ displace the point of the contact element in the same direction leaves the plane MI_1I_2 invariant.

We shall prove that the conditions 1 to 5 completely define the Weyl space.

To prove this, it is sufficient to determine the seven Pfaffian forms in (2). In the meantime we can still normalize the frame (with respect to the coordinates x, y, z, t) by making use of the transformation (21). Before doing this, we remark that the property 3 is expressed analytically by the fact that ∇M is equal to a multiple of I_1, i.e., by the equations

$$\begin{cases} \omega_2' + [\tau\omega_1] + [\pi_2\omega_2] + [\pi_3\omega_3] = 0, \\ \omega_3' + [\tau\omega_2] - [\pi_1\omega_3] + 2[\pi_2\omega_3] = 0. \end{cases} \tag{24}$$

When these relations are satisfied, the property 4 is characterized by the condition that $\Omega_1 = 0$ when (23) holds, i.e., by

$$\omega_1' + [\pi_1\omega_1] + [\pi_3\omega_2] + a[\tau\omega_3] = 0, \tag{25}$$

where a remains undetermined. Similarly, the property 5 is given by an equation of the form

$$\tau' - [\pi_1\tau] + [\pi_2\tau] - b[\omega_1\omega_3] - c[\omega_2\omega_3] = 0. \tag{26}$$

From the second equation of (24) we find, by making use of (20),

$$[(\tau + \theta_1)(p_t dx + q_t dy)] \equiv 0, \qquad \text{mod } \theta.$$

Since τ is a linear combination of θ and θ_1, it follows that τ is of the form

$$\tau = -\theta_1 + \zeta\theta, \tag{27}$$

where ζ is undetermined. We calculate next the first equation of (24), mod ω_2, ω_3. We find then

$$[(\omega_1 + p_{tt}dx + q_{tt}dy)\theta_1] \equiv 0, \qquad \text{mod } \omega_2, \omega_3.$$

But ω_1 is a linear combination of dx, dy, θ only, so that it must be of the form

$$\omega_1 = -(p_{tt}dx + q_{tt}dy) + \sigma(p_t dx + q_t dy) + \xi\theta.$$

By applying the change of frame (21), we find that ω_1 is then changed into $\omega_1 + v\omega_2 + \frac{1}{2}v^2\theta$. We may therefore assume the frame so chosen that

$$\omega_1 = -(p_{tt}dx + q_{tt}dy) + \xi\theta. \tag{28}$$

When ω_1, ω_2, ω_3 are of the forms given by (20), (28), the frame attached to each contact element is uniquely determined.

It now remains to determine the functions ξ, η, ζ and the Pfaffian forms π_1, π_2, π_3 from the conditions (24), (25), (26) and to show that they are thus completely determined. For this purpose we have to calculate the exterior derivatives of ω_1, ω_2, ω_3, τ. We remark that the calculation of every such

exterior derivative is equivalent to the calculation of the exterior derivative of a Pfaffian form of the form

$$\omega = l\theta - mdx - ndy, \tag{29}$$

where l, m, n are functions of x, y, z, t. We find, by making use of (20), (27), (28),

$$\begin{cases} P\omega' = -\left(\dfrac{dm}{dy} - \dfrac{dn}{dx}\right)[\omega_1\omega_2] + \left\{\left(\dfrac{dm}{dy} - \dfrac{dn}{dx}\right)\eta + (m_t q_t - n_t p_t)\zeta \right. \\ \quad \left. + \left(\dfrac{dl}{dx}q_t - \dfrac{dl}{dy}p_t\right) + (m_z q_t - n_z p_t) + l(p_z q_t - q_z p_t)\right\}[\omega_1\omega_3] \\ \quad + \left\{-\left(\dfrac{dm}{dy} - \dfrac{dn}{dx}\right)\xi + (-m_t q_u + n_t p_u - lP)\zeta + \left(-\dfrac{dl}{dx}q_u + \dfrac{dl}{dy}p_u\right)\right. \\ \quad \left. + (-m_z q_u + n_z p_u) + l(-p_z q_u + q_z p_u)\right\}[\omega_2\omega_3] \\ \quad + (n_t p_t - m_t q_t)[\omega_1\tau] + (lP + m_t q_u - n_t p_u)[\omega_2\tau] \\ \quad + \{(m_t q_t - n_t p_t)\xi + (-m_t q_u + n_t p_u - lP)\eta + l_t P\}[\omega_3\tau]. \end{cases} \tag{30}$$

On substituting into (24), (25), (26), for ω_1', ω_2', ω_3', τ their expressions in terms of the exterior products of ω_1, ω_2, ω_3, τ, we shall get relations of the form

$$\begin{cases} -[\pi_1\omega_1] - [\pi_3\omega_2] = \lambda_1[\omega_1\omega_2] + \lambda_2[\omega_1\omega_3] + \lambda_3[\omega_2\omega_3] + \lambda_4[\omega_1\tau] + \lambda_5[\omega_2\tau], \\ -[\pi_2\omega_2] - [\pi_3\omega_3] = \mu_1[\omega_1\omega_2] + \mu_2[\omega_1\omega_3] + \mu_3[\omega_2\omega_3] + \mu_4[\omega_2\tau] + \mu_5[\omega_3\tau], \\ [\pi_1\omega_3] - 2[\pi_2\omega_3] = \nu_1[\omega_1\omega_3] + \nu_2[\omega_2\omega_3] + \nu_3[\omega_3\tau], \\ [\pi_1\tau] - [\pi_2\tau] = \rho_1[\omega_1\tau\ ] + \rho_2[\omega_2\tau] + \rho_3[\omega_3\tau], \end{cases} \tag{31}$$

where we have, in particular,

$$\begin{cases} \lambda_1 = \dfrac{1}{P}\left(\dfrac{dq_u}{dx} - \dfrac{dP_u}{dy}\right), \qquad \lambda_4 = \dfrac{1}{P}(p_t q_{uu} - q_t p_{uu}), \\ \qquad\qquad \lambda_5 = \xi - \dfrac{1}{P}(p_u q_{uu} - q_u p_{uu}), \\ \mu_1 = \dfrac{1}{P}\left(\dfrac{dq_t}{dx} - \dfrac{dp_t}{dy}\right), \qquad \mu_4 = \eta, \qquad \mu_5 = -\xi - \eta^2 + \eta_t, \\ \mu_2 = \dfrac{1}{P}\left(\dfrac{d\eta}{dx}q_t - \dfrac{d\eta}{dy}p_t\right) + \dfrac{\eta}{P}\left(\dfrac{dp_t}{dy} - \dfrac{dq_t}{dx} + p_z q_t - q_z p_t\right) - \zeta \\ \qquad\qquad + \dfrac{1}{P}(-p_t q_{tz} + q_t p_{tz}), \\ \nu_1 = \dfrac{1}{P}(p_z q_t - q_z p_t), \qquad \nu_2 = \dfrac{1}{P}(-p_z q_u + q_z p_u) - \zeta, \qquad \nu_3 = -\eta, \\ \rho_1 = \dfrac{1}{P}(r_t q_t - s_t p_t), \qquad \rho_2 = \zeta + \dfrac{1}{P}(-r_t q_u + s_t p_u). \end{cases} \tag{32}$$

In order that the system of equations (31) have a set of solutions for π_1, π_2, π_3 it is necessary and sufficient that the following relations hold:

$$(33)\qquad \begin{cases} \lambda_1 - \mu_2 + \nu_2 - 2\rho_2 = 0, & \lambda_4 + \nu_3 - 2\mu_4 = 0, \\ \lambda_5 - \mu_5 = 0, & \mu_1 - \nu_1 + \rho_1 = 0. \end{cases}$$

The set of solutions is then given by

$$(34)\qquad \begin{cases} \pi_1 = (-\nu_1 + 2\rho_1)\omega_1 + (-\nu_2 + 2\rho_2)\omega_2 + \lambda_2\omega_3 + (-\nu_3 + 2\mu_4)\tau, \\ \pi_2 = (-\nu_1 + \rho_1)\omega_1 + (-\nu_2 + \rho_2)\omega_2 + (\lambda_2 - \rho_3)\omega_3 + \mu_4\tau, \\ \pi_3 = -\mu_2\omega_1 + (\lambda_2 - \mu_3 - \rho_3)\omega_2 + \lambda_3\omega_3 + \mu_5\tau, \end{cases}$$

and is uniquely determined. Of the relations (33) the last one is identically satisfied, while the other three equations determine ξ, η, ζ. The Weyl connection is thus completely determined.

On summing up our results, we get the theorem:

Suppose that there is given in the space (x, y, z) a two-parameter family of surfaces such that the tangent planes to the surfaces through a point do not pass through a fixed direction. Then there can be defined in the space, in an intrinsic way, one and only one Weyl connection of elements having the following properties:

1. *Points are developed into points.*
2. *Surfaces of the family are developed into planes.*
3. *The infinitesimal transformation associated to every cycle displaces the point of the element in the direction of the characteristic of the element.*
4. *The infinitesimal transformation associated to every cycle by which the directions d and δ displace the surface of the element in the same "direction" leaves the element invariant.*
5. *The infinitesimal transformation associated to every cycle by which the directions d and δ displace the origin of the element in the same direction leaves the plane of the element invariant.*

3. The Problem of Equivalence

In the study of the geometry of a two-parameter family of surfaces in space under the group of transformations (9) the fundamental problem is the following: Given a two-parameter family $\{\Sigma\}$ of surfaces in S and another such family $\{\bar{\Sigma}\}$ in a space $\bar{S}$ with the coordinates $\bar{x}, \bar{y}, \bar{z}$. When are they *equivalent*, i.e., when can the one be carried into the other by a transformation (9)? We shall show in this section that the solution of this problem follows as a consequence of the theorem of the last section, e.g., the possibility of defining an intrinsic Weyl connection in the space.

Before discussing this so-called "problem of equivalence," we shall give some further properties of the Weyl geometry defined in the last section. Our choice of the frames in §2 adapted to each contact element made use of the coordinates of the space. If we suppose only that the point M of the frame $MI_1I_2I_3$ coin-

cides with the point of the contact element and that the plane MI_1I_2 with the plane of the element, the frame attached to each contact element will depend on three parameters and the relation between two frames attached to the same element will be given by (19). By taking the parameters u, v, w in (19) as independent variables, we get a family of frames depending on the seven parameters x, y, z, t, u, v, w. Then the seven Pfaffian forms ω_1, ω_2, ω_3, τ, π_1, π_2, π_3 are linearly independent and it is easy to verify that they satisfy the system of equations (5), (6), which are called *the equations of structure of the Weyl space*. For this general family of frames the conditions

$$\Omega_1 = a[\omega_3\tau], \qquad \Omega_2 = \Omega_3 = 0, \qquad \mathrm{T} = b[\omega_1\omega_3] + c[\omega_2\omega_3] \tag{35}$$

are still satisfied, because they have an intrinsic geometric meaning.

On account of these forms of the expressions for Ω_1, Ω_2, Ω_3, T the expressions for Π_1, Π_2, Π_3 can be simplified. In fact, by applying to the first four equations of (5) the theorem that the exterior derivative of the exterior derivative of a Pfaffian form is zero (the so-called "Poincaré's Theorem") and noticing that the resulting equations are identically satisfied if the expressions in (6) are zero, we get

$$\begin{cases} \Omega_1' = [\Pi_1\omega_1] + [\Pi_3\omega_2] - [\Omega_1\pi_1] - [\Omega_2\pi_3], \\ \Omega_2' = [\mathrm{T}\omega_1] + [\Pi_2\omega_2] + [\Pi_3\omega_3] - [\Omega_1\tau] - [\Omega_2\pi_2] - [\Omega_3\pi_3], \\ \Omega_3' = [\mathrm{T}\omega_2] - [\Pi_1\omega_3] + 2[\Pi_2\omega_3] - [\Omega_2\tau] + [\Omega_3\pi_1] - 2[\Omega_3\pi_2], \\ \mathrm{T}' = [\mathrm{T}\pi_1] - [\mathrm{T}\pi_2] - [\Pi_1\tau] + [\Pi_2\tau]. \end{cases} \tag{36}$$

These are some of the "Bianchi Identities" of the Weyl space. When the conditions (35) are satisfied, these equations can be simplified to the form

$$\begin{cases} [\Pi_1\omega_1] + [\Pi_3\omega_2] - [(da + 3a\pi_1 - 3a\pi_2)\omega_3\tau] = 0, \\ [\Pi_2\omega_2] + [\Pi_3\omega_3] + c[\omega_1\omega_2\omega_3] = 0, \\ [\Pi_1\omega_3] - 2[\Pi_2\omega_3] + b[\omega_1\omega_2\omega_3] = 0, \\ [\Pi_1\tau] - [\Pi_2\tau] - b[\omega_1\omega_2\tau] + [(b\omega_1 + c\omega_2)'\omega_3] + [(b\omega_1 + c\omega_2)\pi_2\omega_3] = 0. \end{cases} \tag{37}$$

From the third equation we see that every term of $\Pi_1 - 2\Pi_2 + b[\omega_1\omega_2]$ must contain ω_3. On the other hand, we get, by multiplying the fourth equation by ω_3,

$$[(\Pi_1 - \Pi_2 - b\omega_1\omega_2)\tau\omega_3] = 0.$$

Similarly, the second and the third equations give respectively

$$[\Pi_2\omega_2\omega_3] = 0, \qquad [(\Pi_1 - 2\Pi_2)\omega_2\omega_3] = 0,$$

from which it follows that

$$[(\Pi_1 - \Pi_2 - b\omega_1\omega_2)\omega_2\omega_3] = 0.$$

Again, by multiplying the right-hand sides of the first equation by ω_3, the second by ω_2, the third by $-\omega_1$, and adding, we get

$$[(\Pi_1 - \Pi_2 - b\omega_1\omega_2)\omega_1\omega_3] = 0.$$

It follows from the three equations that every term of $\Pi_1 - \Pi_2 - b[\omega_1\omega_2]$ must contain ω_3. We can therefore put

$$(38) \qquad \begin{cases} \Pi_1 = 3b[\omega_1\omega_2] + e[\omega_1\omega_3] + f[\omega_2\omega_3] + g[\omega_3\tau], \\ \Pi_2 = 2b[\omega_1\omega_2] + h[\omega_1\omega_3] + i[\omega_2\omega_3] + j[\omega_3\tau]. \end{cases}$$

From (37) it then follows that Π_3 is of the form

$$(39) \qquad \Pi_3 = (h - c)[\omega_1\omega_2] + f[\omega_1\omega_3] + j[\omega_2\tau] + k[\omega_2\omega_3] + l[\omega_3\tau].$$

Consequently, for our Weyl space, the seven Pfaffian forms ω_1, ω_2, ω_3, τ, π_1, π_2, π_3 satisfy the equations of structure (5), with Ω_1, Ω_2, Ω_3, T, Π_1, Π_2, Π_3 given by (35), (38), (39). In these equations there are introduced 11 quantities $a, b, c, e, f, g, h, i, j, k, l$, which are functions of x, y, z, t, u, v, w and which constitute the tensor of curvature and torsion of the space.

When two such Weyl spaces of elements are given, with the coordinates (x, y, z, t) and $(\bar{x}, \bar{y}, \bar{z}, \bar{t})$, we say that they are equivalent, if to every family of frames of the one (with one frame attached to each contact element), there is a corresponding family of frames of the other, such that a transformation of the form (7) exists, which satisfies the relations

$$(40) \qquad \bar{\omega}_i = \omega_i, \qquad \bar{\tau} = \tau, \qquad \bar{\pi}_i = \pi_i, \qquad i = 1, 2, 3.$$

If we attach to each contact element (x, y, z, t) the most general family of frames depending on three parameters u, v, w and to $(\bar{x}, \bar{y}, \bar{z}, \bar{t})$ the family of frames with the parameters $\bar{u}, \bar{v}, \bar{w}$, it is evident that the two Weyl spaces are equivalent, when and only when there exists a transformation of the form

$$(41) \qquad \begin{cases} \bar{x} = \bar{x}(x, y, z, t, u, v, w), \cdots, & \bar{t} = \bar{t}(x, y, z, t, u, v, w), \\ \bar{u} = \bar{u}(x, y, z, t, u, v, w), \cdots, & \bar{w} = \bar{w}(x, y, z, t, u, v, w), \end{cases}$$

satisfying the relations (40).

With these preparations we now establish the theorem:

There exists a point transformation of the form (9) *carrying one family of surfaces* $\{\Sigma\}$ *into another* $\{\bar{\Sigma}\}$ *when and only when the two corresponding Weyl spaces of elements are equivalent.*

The proof of this theorem is immediate. In fact, when the two families of surfaces are equivalent, the corresponding Weyl spaces are equivalent, since our definition of the Weyl space is intrinsic. Conversely, when the Weyl spaces are equivalent, the transformation (41) will realize the equality of the sets of Pfaffian forms in (40). From the first three equations in (40), we see that in

the transformation (41) the coordinates $\bar{x}$, $\bar{y}$, $\bar{z}$ are functions of x, y, z alone, so that the transformation in question is a point transformation. From

$$\bar{\omega}_3 = \omega_3 , \qquad \bar{\tau} = \tau$$

we conclude that the transformation carries $\{\Sigma\}$ to $\{\bar{\Sigma}\}$. Hence the theorem is proved.

Analytically this result may be formulated as follows: *To solve the problem of equivalence for a two-parameter family of surfaces (whose tangent planes at a point do not pass through a fixed direction) in a three-dimensional space (x, y, z), we introduce four auxiliary variables t, u, v, w and seven linearly independent Pfaffian forms ω_i, τ, π_i $(i = 1, 2, 3)$ in the coordinates and in these variables. All these Pfaffian forms are invariant in the sense that the two families of surfaces $\{\Sigma\}$ and $\{\bar{\Sigma}\}$ are equivalent, when and only when there exists a transformation of the form* (41) *satisfying* (40).

Since the process of exterior differentiation is invariant under the point transformations (41), it follows that the 11 quantities $a, \cdots, l$ in (35), (38), (39) are invariants. We shall call them the *fundamental invariants* of the family $\{\Sigma\}$. From a given invariant F the equation

$$dF = F_1\omega_1 + F_2\omega_2 + F_3\omega_3 + F_0\tau + F_1'\pi_1 + F_2'\pi_2 + F_3'\pi_3 \tag{42}$$

defines the new invariants F_1, F_2, F_3, F_0, F_1', F_2', F_3', called the *covariant derivatives* of F. According to a theorem of E. Cartan,[7] *the complete set of invariants of the family of surfaces consists of the fundamental invariants and their successive covariant derivatives.* This is to be understood as follows: By eliminating the independent variables x, y, z, t, u, v, w, we may set up relations of the form

$$\Phi(F_1, \cdots, F_m) = 0 \tag{43}$$

between the invariants $F_1, \cdots, F_m$ of our complete set. *Then the two families of surfaces $\{\Sigma\}$ and $\{\bar{\Sigma}\}$ are equivalent if and only if the relations* (43) *and*

$$\Phi(\bar{F}_1, \cdots, \bar{F}_m) = 0$$

hold at the same time for $F_1, \cdots, F_m$ and for the corresponding invariants $\bar{F}_1, \cdots, \bar{F}_m$ of $\{\bar{\Sigma}\}$. Moreover, it is also well-known that the equivalence of $\{\Sigma\}$ and $\{\bar{\Sigma}\}$ can be decided by only a finite number of such relations (43).

4. The Weyl Space of Points

An important particular case of our geometry is the case by which the infinitesimal transformation associated to every cycle depends only on the displacements of the origins of the contact elements. The space of elements is

[7] E. Cartan, *Les sousgroupes des groupes continus de transformations*, Annales de l'Ecole Normale Supérieure, série 3, t. 25 (1908), pp. 57–194.

then identical with the three-dimensional point space (x, y, z) with a Weyl connection. The conditions for this are, from (35), (38), (39),

$$a = g = j = l = 0.$$

These conditions are, however, not independent. In fact, when $a = 0$, the first equation of (37) will give

$$g = l = 0.$$

It follows that *the conditions*

$$a = j = 0 \tag{44}$$

are the necessary and sufficient conditions for the Weyl space to be a Weyl space of points.

We may give a geometrical interpretation to the condition $a = 0$. In fact, *it signifies that the tangent planes to the surfaces of the family through a fixed point in space envelop a cone of the second order.* To prove this, let δ be an operation under which the point remains fixed, so that

$$\omega_1(\delta) = \omega_2(\delta) = \omega_3(\delta) = 0.$$

Since t is now the only variable, we may determine the auxiliary variables u, v, w as functions of t such that the conditions

$$\pi_1(\delta) = \pi_2(\delta) = \pi_3(\delta) = 0$$

are also satisfied. We put $\tau(\delta) = e$, and we can assume, by changing the parameter t when necessary, that $\delta e = 0$.

If δ is an operation so chosen and d is an operation by which all the variables vary, we get, from (5) and (35),

$$\begin{cases} \delta\omega_1 = -ae\omega_3 , \\ \delta\omega_2 = -e\omega_1 , \\ \delta\omega_3 = -e\omega_2 . \end{cases} \tag{45}$$

From (45) we get

$$\begin{cases} \delta^2\omega_3 = e^2\omega_1 , \\ \delta^3\omega_3 = -ae^3\omega_3 , \\ \delta^4\omega_3 = ae^4\omega_2 + (\cdots)\omega_3 , \\ \delta^5\omega_3 = -ae^5\omega_1 + (\cdots)\omega_2 + (\cdots)\omega_3 . \end{cases} \tag{46}$$

If we define the "plane coordinates" l_1 , l_2 , l_3 of a plane through the point by the relation

$$l_1\omega_1 + l_2\omega_2 + l_3\omega_3 = 0, \tag{47}$$

we shall get, when the coordinates l_1, l_2, l_3 of the tangent plane of the family are expanded in powers of e,

$$\begin{cases} l_1 = \frac{1}{2}e^2 - \frac{1}{120}ae^5 + \cdots, \\ l_2 = -e + \frac{1}{24}ae^4 + \cdots, \\ l_3 = 1 - \frac{1}{6}ae^3 + \cdots. \end{cases} \tag{48}$$

From (48) it follows that

$$l_2^2 - 2l_1l_3 = \frac{1}{10}ae^5 + \cdots. \tag{49}$$

Hence the cone

$$l_2^2 - 2l_1l_3 = 0 \tag{50}$$

is the osculating quadric cone of the cone enveloped by the tangent planes of the family $\{\Sigma\}$ and the condition $a = 0$ is a necessary and sufficient condition that these tangent planes themselves envelop a quadric cone.

Added October 10, 1941. The author has succeeded in extending the results of this paper to a space of $n(\geqq 3)$ dimensions. The following theorem has been established: *Given in a space of n dimensions a family of hypersurfaces depending on $n - 1$ parameters such that the tangent hyperplanes at a point to the hypersurfaces of the family through the point envelop a hypercone of $n - 1$ dimensions. Then it is possible to define in the space, in an intrinsic way, a Weyl geometry.* The elements of this Weyl space are in general more complicated for $n \geqq 4$ than for $n = 3$. The problem of equivalence is solved as a consequence of this result.

National Tsing Hua University
Kunming, China

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

IDEMPOTENT MARKOFF CHAINS

BY DAVID BLACKWELL

(Received March 6, 1942)

I. Introduction

Let $\mathfrak{B}$ be any Borel field of subsets of an abstract space X, and suppose that for each $x \in X$ a probability measure[1] $P(x, E)$ is defined on $\mathfrak{B}$ and that $P(x, E)$ is for fixed E a $\mathfrak{B}$-measurable function of x. Then $P(x, E)$ may be considered as representing the transition probability of going from the point x into the set E in a single trial, and it is said to determine a *Markoff chain* on X. The probability of going from x into E in n trials, denoted by $P_n(x, E)$, is given inductively by

$$P_n(x, E) = \int P(y, E)\, dP_{n-1}(x, y).^2 \tag{1}$$

In this paper we shall consider Markoff chains for which $P_n(x, E)$ is independent of n. It is clear from (1) that this will occur whenever $P_2(x, E) \equiv P(x, E)$, so that we shall be studying Markoff chains which satisfy

$$P(x, E) = \int P(y, E)\, dP(x, y). \tag{2}$$

Such a Markoff chain will be called *idempotent*, and the justification for this is apparent.

Besides having some independent interest, idempotent Markoff chains are useful as tools in the study of general Markoff chains. For example if there is a subsequence $N_\nu \to \infty$ such that for all x and E

$$\frac{1}{N_\nu} \sum_{n=1}^{N_\nu} P_n(x, E) \to Q(x, E),$$

where $Q(x, E)$ is for each x a probability measure, then the Markoff chain determined by $Q(x, E)$ is idempotent, and the properties of $Q(x, E)$ are closely related to those of $P(x, E)$ (cf. Doob (II) and Yosida and Kakutani (III)).

Suppose that X contains only a finite number of points, denoted by the numbers $1, \cdots, n$, and suppose that $\mathfrak{B}$ consists of all subsets of X. Then $P(x, E)$ may be represented by a *Markoff matrix* $P = \| p_{ij} \|$, i.e. a square matrix with non-negative elements and row sums equal to unity, where p_{ij} denotes the probability of going from i to j in a single trial. For this case (2) is the statement that the matrix P is idempotent. Idempotent Markoff matrices

[1] A probability measure is a non-negative completely additive set function having the value unity for the space.

[2] We shall use the notation $dP_n(x, y)$ to denote integration of y with respect to the measure $P_n(x, E)$.

have been completely characterized by Doob (II) and by Yosida and Kakutani (III), and the result is essentially summarized in the following theorem:

THEOREM 1: *If* $P = \| p_{ij} \|$ *is an idempotent Markoff matrix, the subscripts may be divided into groups* $F, A_1, \cdots, A_k$ *such that*

(a) $$p_{ij} = p_j > 0 \quad \text{for} \quad i, j \in A_l$$

(b) $$\sum_{j \in A_l} p_j = 1$$

(c) $$p_{ij} \equiv 0 \quad \text{for} \quad j \in F.$$

In section II we shall prove some general theorems about idempotent Markoff chains. In sections III and IV we shall apply these results to two classes of idempotent Markoff chains and show that in each case a decomposition analogous to that described by Theorem 1 obtains. Finally in section V we give a simple example of an idempotent Markoff chain which admits no such decomposition.

II. Definitions and general theorems

In this section we shall restrict attention to idempotent Markoff chains A set $N \in \mathfrak{B}$ is called a *null set* if $P(x, N) = 0$ for all x. A set $I \in \mathfrak{B}$ is called *invariant* if there is a null set N such that

$$P(x, I) = 1 \quad \text{for all } x \in I - N.$$

The equation

(3) $$\int_E P(y, CE)\, dP(x, y) = 0 \quad \text{for all } x$$

is clearly a necessary and sufficient condition for the invariance of E. Following Doeblin (I), we shall call an invariant set *indecomposable* if it does not contain two disjoint non-null invariant subsets. A set E is called strictly *invariant* if $P(x, E) = 1$ for all $x \in E$. It is clear that any strictly invariant set is invariant and that if E is invariant there is a null set N such that $E - N$ is strictly invariant. We shall need the following fact about the reduction of multiple integrals:

LEMMA 1: *Let* $\mathfrak{B}(\mathfrak{B}')$ *be a Borel field of subsets of a space* $Z(Y)$, *and suppose that* $P(y, E)$ *is for each* $y \in Y$ *a measure on* $\mathfrak{B}$ *and for fixed* E *a* $\mathfrak{B}'$*-measurable function of* y. *Let* m *be any measure on* $\mathfrak{B}'$, *and define*

$$M(E) = \int P(y, E)\, dm.$$

Then any M*-integrable function* $f(z)$ *is integrable with respect to* $P(y, E)$ *for all* y *except a set of* m*-measure zero, and*

$$\int f(z)\, dM = \int \left\{ \int f(z)\, dP(y, z) \right\} dm.$$

PROOF: The verification of the lemma is immediate when $f(z)$ is the characteristic function of a set of finite M-measure, and the general result follows by approximation.

THEOREM 2: *If S, E are any sets, then for all x*

$$\int_S \left\{ \int_{CS} P(z, E)\, dP(y, z) \right\} dP(x, y) = \int_{CS} \left\{ \int_S P(z, E)\, dP(y, z) \right\} dP(x, y). \tag{4}$$

This theorem states that, starting from any point x, the probability of going first into S, then into CS, and finally into E is the same as the probability of going first into CS, then into S, and finally into E. This is not of course true for general Markoff chains, and in fact it may be shown that if it is true for every S in the special case $E = X$, then the Markoff chain is idempotent. The idea of the proof to be given below and of the results which follow will become clear at once if the integrals are interpreted as representing probabilities of events.

PROOF: For all x,

$$\begin{aligned} &\int_S \left\{ \int_{CS} P(z, E)\, dP(y, z) \right\} dP(x, y) + \int_{CS} \left\{ \int_{CS} P(z, E)\, dP(y, z) \right\} dP(x, y) \\ &\qquad = \int \left\{ \int F(z) P(z, E)\, dP(y, z) \right\} dP(x, y), \end{aligned} \tag{5}$$

where $F(z)$ is the characteristic function of CS. Also

$$\begin{aligned} &\int_{CS} \left\{ \int_S P(z, E)\, dP(y, z) \right\} dP(x, y) + \int_{CS} \left\{ \int_{CS} P(z, E)\, dP(y, z) \right\} dP(x, y) \\ &\qquad = \int_{CS} \left\{ \int P(z, E)\, dP(y, z) \right\} dP(x, y) = \int F(y) P(y, E)\, dP(x, y). \end{aligned} \tag{6}$$

By Lemma 1, with $f(z) = F(z)P(z, E)$, the right members of (5) and (6) are equal. Hence their left members are equal, and (4) follows at once.

THEOREM 3: *The class $\mathcal{I}$ of invariant sets is a Borel field.*

PROOF: It is clear that a denumerable sum of invariant sets is invariant; we need show only that the complement of an invariant set is also invariant. Taking $E = X$ in Theorem 2, (4) becomes

$$\int_S P(y, CS)\, dP(x, y) = \int_{CS} P(y, S)\, dP(x, y). \tag{7}$$

If S is invariant, it follows immediately from this equation and condition (3) that CS is also invariant.

THEOREM 4: *The function $P(x, E)$ is for fixed E an $\mathcal{I}$-measurable function.*

PROOF: We must show that the set

$$S = \{x \text{ for which } P(x, E) < k\}$$

is invariant for any k. We may assume $k > 0$. Now

$$\int_{CS}\left\{\int_S P(z, E)\, dP(y, z)\right\} dP(x, y) \leqq k \int_{CS} P(y, S)\, dP(x, y), \tag{8}$$

and the equality sign holds only if both sides vanish. Also

$$\int_S\left\{\int_{CS} P(z, E)\, dP(y, z)\right\} dP(x, y) \geqq k \int_S P(y, CS)\, dP(x, y). \tag{9}$$

By Theorem 2 the left members of (8) and (9) are equal, and (7) implies that their right members are equal. Hence equality holds in (8) for all x. Consequently the right side of (8) vanishes for all x and, since $k > 0$, this is precisely condition (3) for invariance of CS.

The principal result of this section is contained in the following theorem which shows that the strictly invariant indecomposable sets are the analogues of the sets A_l of Theorem 1.

THEOREM 5: *If S is strictly invariant and indecomposable, then $P(x, E)$ is independent of x for all $x \in S$.*

PROOF: For all $z \in S$ we have

$$P(z, E) = \int_S P(y, E)\, dP(z, y). \tag{10}$$

Let $y_0 \in S$. The sets

$$S\cdot\{P(x, E) \geqq P(y_0, E)\} \quad \text{and} \quad S\cdot\{P(x, E) \leqq P(y_0, E)\}$$

are by Theorem 4 invariant, and it is clear from (10) with $z = y_0$ that neither can be a null set. Since S is indecomposable, their complements in S, the sets

$$S\cdot\{P(x, E) < P(y_0, E)\} \quad \text{and} \quad S\cdot\{P(x, E) > P(y_0, E)\}$$

must be null sets. Then $P(x, E) = P(y_0, E)$ for all $x \in S$ except a null set, and (10) then implies that

$$P(x, E) = P(y_0, E) \quad \text{for all } x \in S.$$

III. The strictly separable case

A Borel field $\mathfrak{B}$ of subsets of X is said to be *strictly separable* if it is determined by a denumerable subcollection of its elements. $\mathfrak{B}$ is said to be *atomic* if $X = \sum_\alpha X_\alpha$, where each $X_\alpha \in \mathfrak{B}$ and (11) for every $B \in \mathfrak{B}$ and every α, either $BX_\alpha = X_\alpha$ or $BX_\alpha = O$. The sets X_α are called the *atoms* of $\mathfrak{B}$.

LEMMA 2: *A strictly separable Borel field is atomic.*

This fact is well known; since we shall refer to the proof, we reproduce it here.

PROOF: Let $\mathfrak{B}$ be determined by $F_1, F_2, \cdots$. Define $X_\alpha = \prod_{n=1}^{\infty} F_n^{\epsilon_n(\alpha)}$, where $\epsilon_n(\alpha) = \pm 1$, $E^1 = E$, $E^{-1} = CE$. Then each $X_\alpha \in \mathfrak{B}$, $X = \sum_\alpha X_\alpha$, and the class of sets B satisfying (11) is a Borel field including every F_n and hence includes $\mathfrak{B}$.

THEOREM 6: *If $\mathcal{G}$ contains a strictly separable subfield $\mathcal{G}_1$ whose atoms are indecomposable, then $X = F + \sum_\alpha A_\alpha$, where*

(a) $$P(x, E) = P_\alpha(E) \text{ for } x \in A_\alpha$$

(b) $$P_\alpha(A_\alpha) = 1$$

(c) *F is a null set.*

PROOF: Let $F_1, F_2, \cdots$ determine $\mathcal{G}_1$. We may assume the F_k form a field. Let N_k be the null subset of F_k for which $P(x, F_k) < 1$, and define $F = \sum_1^\infty N_k$. Let $\{X_\alpha\}$ be the atoms of I, and define $A_\alpha = X_\alpha - F$. If $x \in A_\alpha$, it is evident from the construction of the atoms given in the proof of Lemma 2 that $P(x, X_\alpha) = 1$ and hence that $P(x, A_\alpha) = 1$. Since A_α is strictly invariant and indecomposable, we may invoke Theorem 5 to obtain (a).

THEOREM 7: *If $\mathcal{B}$ is strictly separable, the conclusion of Theorem 6 holds.*

PROOF: We shall define a subfield of $\mathcal{G}$ satisfying the hypothesis of Theorem 6. Let $F_1, F_2, \cdots$ determine $\mathcal{B}$. We may suppose the F_j form a field. Define

$$I_{jkn} = \left\{\frac{k}{n} \leqq P(x, F_j) < \frac{k+1}{n}\right\}, \qquad k = 0, \cdots, n;\ n, j = 1, 2, \cdots.$$

Each of these sets is invariant by Theorem 4. Let $\mathcal{G}_1$ be the strictly separable subfield determined by the I_{jkn}. If X_α is an atom of $\mathcal{G}_1$, it is clear that $P(x, F_j)$ is independent of x for all $x \in X_\alpha$ and all j. Since whenever a set of measures coincide on a field, they coincide on the Borel field determined by it, $P(x, E)$ is independent of x for all $x \in X_\alpha$ and all E. Then each X_α is indecomposable, for otherwise it would contain two disjunct non-empty strictly invariant subsets S and T, and $P(x, S) = 1$ for $x \in S$ and 0 for $x \in T$, contradicting the fact that $P(x, E)$ is independent of x on X_α.

Thus for instance if $\mathcal{B}$ is the class of Borel sets in Euclidean space, any idempotent Markoff chain admits the decomposition of Theorem 6.

IV. The density case

Let m be any finite measure defined on $\mathcal{B}$ and define the product measure $m \times m$ on the Borel field $\mathcal{B}_1 = \mathcal{B} \times \mathcal{B}$ of subsets of the product space $X \times Y$ of X with itself. Suppose $p_1(x, y)$ is any non-negative $\mathcal{B}_1$-measurable function such that for each $x \in X$,

(12) $$\int p_1(x, y)\, dy = 1.$$

Then the function

(13) $$P(x, E) = \int_E p_1(x, y)\, dy$$

determines a Markoff chain on X, and the function $p_1(x, y)$ is called the *density function*. If the Markoff chain so defined is idempotent, we have

$$\int_E p_1(x, y)\, dy = \int \left\{ \int_E p_1(z, y)\, dy \right\} dP(x, z). \tag{14}$$

Applying a well known formula for transformation of measures and interchanging the order of integration, we obtain

$$\int_E p_1(x, y)\, dy = \int_E \left\{ \int p_1(x, z) p_1(z, y)\, dz \right\} dy; \tag{15}$$

i.e. if we define $p(x, y)$ by the equation

$$p(x, y) = \int p_1(x, z) p_1(z, y)\, dz, \tag{16}$$

$p(x, y)$ is another density function defining exactly the same Markoff chain as $p_1(x,y)$. It follows that for fixed x, $p(x, z)$ and $p_1(x, z)$ are equal for almost all z, and (16) then implies

$$p(x, y) = \int p(x, z) p_1(z, y)\, dz. \tag{17}$$

Multiplying both sides of (17) by $p_1(y, w)$ and integrating with respect to y, we obtain

$$p(x, w) = \int p(x, z) p(z, w)\, dz. \tag{18}$$

Thus if an idempotent Markoff chain can be represented by a density function, it can be represented by a density function which is idempotent in the sense of (18). (18) is the direct analogue of the idempotent Markoff matrix if we think of the Markoff chain as determined by the density function $p(x, y)$ rather than by the function $P(x, E)$. To study idempotent density functions we shall use the following well known fact:

LEMMA 3: *If m is a finite measure defined on a Borel field $\mathfrak{B}$ of subsets of a space X, then*

$$X = S + X_1 + X_2 + \cdots \quad \textit{where} \tag{19}$$

(a) $m(X_i) > 0$, *and* $E \in \mathfrak{B}$, $E \subseteq X_i$ *implies* $m(E) = m(X_i)$ *or* $m(E) = 0$.
(b) *every subset of S of positive measure has a subset of every smaller positive measure.*

Either S or the X_i may of course be absent. The decomposition is clearly unique up to sets of measure zero.

PROOF: The sets X_i satisfying (a) may be divided into equivalence classes by defining two such sets as equivalent if their symmetric difference has measure zero. Any two sets in different classes can have only a set of measure zero in common, and since each set has positive measure there are at most denumerably

many equivalence classes. Let $X_1, X_2, \cdots$ be a set of representatives, one from each equivalence class, and define $S = X - \sum_1^\infty X_n$. Then every subset of S of positive measure has a subset of smaller positive measure and hence a subset of arbitrarily small positive measure, and we must show that this implies (b). Suppose $S_1 \subseteq S, 0 < k < m(S_1)$. Choose a sequence of sets K_m inductively so as to satisfy

(c) $$K_n \in \mathcal{R}_n$$

(d) $$m(K_n) > \text{l.u.b.}_{E \in \mathcal{R}_n} m(E) - \frac{1}{n}$$

where $\mathcal{R}_n$ is the class of all subsets E of $S_1 - (K_1 + \cdots + K_{n-1})$ for which $m(E) \leqq k - m(K_1 + \cdots + K_{n-1})$. Define $K = \sum_1^\infty K_n$. Then $K \subseteq S_1$ and $m(K) = k$, otherwise the choice of K_n would at some stage be contradicted, since $S_1 - K$ would contain a subset of positive measure smaller than $k - m(K)$.

We now use an argument similar to that of Doeblin in a related discussion (I). Apply Lemma 3 to m defined on the Borel field $\mathcal{I}$ of invariant subsets of X, obtaining a decomposition (19) into invariant subsets. Each X_n is clearly indecomposable, and we may suppose that each X_n is strictly invariant. We shall show that S is a null set. First extract from S a null set H of largest possible measure and write $T = S - H$. It is sufficient to show that T has measure zero. Certainly every null subset of T has measure zero. Choose a class $D_{mn} \in \mathcal{I}$, $n = 1, \cdots, 2^m$; $m = 0, 1, 2, \cdots$ of subsets of T so that

(a) $$D_{mi} \cdot D_{mj} = O \quad \text{if} \quad i \neq j$$

(b) $$m(D_{mi}) = \frac{1}{2^m} m(T)$$

(c) $$D_{m+1,2i-1} + D_{m+1,2i} = D_{mi}.$$

By deleting a null set N_{mi} we shall make each of the D_{mi} strictly invariant. Hence by deleting $N = \sum_{m,i} N_{mi}$ from each D_{mi} simultaneously we obtain a new class of sets satisfying (a), (b), (c) and each of which is moreover strictly invariant. Thus without loss of generality we may suppose that each D_{mi} is strictly invariant. Now

$$D_{01} = \prod_{m=1}^{\infty} (D_{m1} + \cdots + D_{m2^m}) = \sum_{i,j,k,\cdots} D_{1i} \cdot D_{2j} \cdot D_{3k} \cdots .$$

Each set $D_{1i} \cdot D_{2j} \cdot D_{3k} \cdot \cdots$ is a strictly invariant set of measure zero and is therefore empty. It follows that D_{01} is empty. Since $m(T) = m(D_{01})$, T has measure zero.

We are now in a position to apply Theorem 6, and we have shown further that there will be only denumerably many sets A_α in this case. However without appealing to this theorem we may easily obtain an even more precise decomposition in terms of the density function $p(x, y)$.

By Theorem 5, $P(x, E)$ is independent of x for $x \in X_n$; it follows that for any $x_1, x_2 \in X_n$,

$$p(x_1, y) = p(x_2, y) \tag{20}$$

except possibly on a y-set of measure zero. (18) then implies that (20) holds identically in y for all $x_1, x_2 \in X_n$. We write for all $x \in X_n$,

$$p(x, y) = p_n(y).$$

Then for any x it follows from (18) that

$$p(x, y) = \sum_{n=1}^{\infty} \int_{X_n} p(x, z) p_n(y)\, dz = \sum_{n=1}^{\infty} P(x, X_n) p_n(y);$$

i.e. any $p(x, y)$ is a linear combination of the functions $p_n(y)$. Define $U_n = \{y \in X_n$ such that $p_n(y) = 0\}$ and $V_n = \{y \in X_n$ such that $p_n(y) > 0\}$ and write $U = \sum_1^\infty U_n$, $V = \sum_1^\infty V_n$. Then $m(V) = 0$ and both U and V are null sets. Writing $A_n = X_n - (U + V)$, $F = X - (\sum_1^\infty X_n + V)$, we may summarize our results in the following theorem.

THEOREM 8: *If $p(x, y)$ is an idempotent density function, then $X = F + V + A_1 + A_2 + \cdots$, where*

(a) $$p(x, y) = p_n(y) \quad \text{for } x \in A_n$$

(b) $$p_n(y) > 0 \quad \text{for } y \in A_n$$

(c) $$\int_{A_n} p_n(y) dy = 1$$

(d) $$p(x, y) \equiv 0 \quad \text{for } y \in F$$

(e) $$m(V) = 0.$$

V. An example

We conclude by giving an example of an idempotent Markoff chain for which no decomposition like that obtained in the preceding sections is possible. Let $\mathfrak{B}$ be any Borel field, and define $P(x, E)$ to be the characteristic function of E. Then every set of $\mathfrak{B}$ is strictly invariant; there are no null sets except the empty set. A decomposition of X would consist in writing $X = \sum_\alpha X_\alpha$ where each X_α is indecomposable; and this is clearly impossible whenever $\mathfrak{B}$ is non-atomic.

INSTITUTE FOR ADVANCED STUDY

BIBLIOGRAPHY

I. DOEBLIN, W., *Chaines simples constantes de Markoff*. Ann. Ec. Norm. (3), LVII., Fasc. 2.

II. DOOB, J. L., *Topics in the theory of Markoff chains*. (To appear in the Trans. Am. Math. Soc.)

III. YOSIDA, K., AND KAKUTANI, S., *Markoff processes with a denumerably infinite number of states*. Jap. Journal of Math., 1939, p. 47.

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

BANACH SPACE METHODS IN TOPOLOGY*

BY SAMUEL EILENBERG

(Received January 22, 1942)

1. Introduction

We shall consider the totality $B(X)$ of all continuous real valued bounded functions f defined on a topological space X.[1] With the usual definition of addition and multiplication by real numbers and with the norm

$$\|f\| = \sup_{x \in X} |f(x)|$$

the set $B(X)$ becomes a Banach space.

Banach[2] has established the interesting and important result that two compact[3] metric spaces X and Y are homeomorphic if and only if the corresponding spaces $B(X)$ and $B(Y)$ are isometric. Stone[4] has shown that the hypothesis that X and Y are metric was superfluous. Both authors arrive at their results by establishing the general form of an isometric mapping of $B(X)$ onto $B(Y)$.

The Banach-Stone theorem implies that if X is compact then the topological structure of X is entirely determined by the metric properties of $B(X)$. The first purpose of this paper is to exhibit this relationship more explicitly. By studying the convex subsets of the surface of the unit sphere in $B(X)$ we define an "ideal space" Ξ and prove that if X is compact, then Ξ and X are homeomorphic. From this result we readily get a proof of the Banach-Stone theorem for compact spaces and also for a class of non-compact spaces including all metric spaces.

Although the space Ξ provides theoretically complete means for translating topological properties of a compact X into metric properties of $B(X)$, it will very often lead from simple topological properties of X to involved metric properties of $B(X)$ with no clear geometric interpretation. This makes it an interesting problem to relate "interesting" topological properties of X with "interesting" metric properties of $B(X)$. In this direction we prove that if $B(X)$ is a direct sum $B_1 + B_2$ of two Banach spaces, then $B_i = B(X_i)$ where X_i is an open and closed subset of X. From this we see that the connectedness properties of X translate into properties of $B(X)$ dealing with direct sums.[5]

---

* Presented to the American Mathematical Society April 12 and May 2, 1941.

[1] A *topological space* is a set X with a family of subsets called *open* such that (a) 0 and X are open, (b) the union of any number of open sets is open, (c) the intersection of two open sets is open. If in addition, (d) every two distinct points of X belong to two disjoint open sets, then X is said to be a *Hausdorff space*.

[2] S. Banach. *Théorie des operations linearies*. Warsaw, 1932, p. 170.

[3] A space X is called *compact* if it is a Hausdorff space and if every covering of X by open sets contains a finite subcovering.

[4] M. H. Stone. *Trans. Amer. Math. Soc.*, 41 (1937), p. 469.

[5] Another result in this direction was recently announced by S. Krein and M. Krein (*C. R. U. R. S. S.*, 27 (1940), pp. 427–430): *If X is completely regular then $B(X)$ is separable and only if X is compact metric.*

$B(X)$ is not only a Banach space but also a linear lattice and a normed ring. From these two points of view $B(X)$ has been thoroughly investigated.[6] In particular, necessary and sufficient conditions are known for a linear lattice or normed ring B in order that $B = B(X)$ for some topological space X. The analogous problem for a Banach space B is as yet unsolved and we hope that the methods developed in this paper will lead towards a solution.

2. Tychonoff's compacting

A Hausdorff space X is called *completely regular* if given a closed set $F \subset X$ and a point $x_0 \in X - F$ there is a continuous real valued function f on X such that $f(x_0) = 0$ and $f(x) = 1$ for all $x \in F$. Corresponding to every completely regular space X Tychonoff[7] has constructed a space $\beta(X)$ such that

(T_1) $\beta(X)$ is compact,

(T_2) $X \subset \beta(X)$, $\bar{X} = \beta(X)$,

(T_3) every $f \in B(X)$ has an extension $f^* \in B[\beta(X)]$.

Moreover, these properties determine $\beta(X)$ uniquely. Since X is dense in $\beta(X)$ it is clear that the extension f^* of $f \in B(X)$ is unique and therefore establishes a $1 - 1$ correspondence between $B(X)$ and $B[\beta(X)]$, in view of which the two Banach spaces may be considered identical.

The space $\beta(X)$ was extensively studied by Čech[8] who has proved among other results that if X is completely regular and satisfies the first countability axiom[9] then $\beta(X)$ determines X. More precisely, X is then the subset of $\beta(X)$ consisting of the points at which $\beta(X)$ has a countable base.[9]

3. Stars

Given a Banach space B we shall consider the transformation $Tb = -b$ and define

$$SA = A \cup TA$$

for every $A \subset B$.

LEMMA 3.1. *Given* $b_1, b_2 \in B$, $\| b_1 \| \leqq 1$, $\| b_2 \| \leqq 1$ *the following conditions are equivalent*:

(a) $\| tb_1 + (1 - t)b_2 \| = 1$ *for at least one* t *such that* $0 < t < 1$,

(b) $\| tb_1 + (1 - t)b_2 \| = 1$ *for all* t *such that* $0 \leqq t \leqq 1$,

(c) $\| b_1 + b_2 \| = 2$.

---

[6] See Stone *loc. cit.* and S. Kakutani, *Ann. of Math.*, 42 (1941), pp. 994–1024. The latter paper contains an up-to-date bibliography of the subject.

[7] A. Tychnoff, *Math. Ann.*, 102 (1930), pp. 544–561.

[8] E. Čech. *Ann. of Math.*, 38 (1937), pp. 823–844.

[9] We say that a sequence U_i of open sets is a *countable base* for the point x of a Hausdorff space X if every open set containing x contains at least one of the sets U_i. If a countable base exists for every point of X, then X is said to satisfy the *first countability axiom*.

The proof is left to the reader.

For every $b \in B$ such that $\| b \| \leqq 1$ we define the star $\mathrm{St}(b)$ as follows:

$$\mathrm{St}(b) = \{b_1 \mid \| b_1 \| \leqq 1, \| b_1 + b \| = 2\}.$$

The definition is justified geometrically by Lemma 3.1. We shall also consider the symmetrized star $\mathrm{SSt}(b)$. Clearly the conditions $\| b \| < 1$, $\mathrm{St}(b) = 0$ and $\mathrm{SSt}(b) = 0$ are equivalent. Also the conditions $b_1 \in \mathrm{St}(b_2)$ and $b_2 \in \mathrm{St}(b_1)$ are equivalent and similarly for the symmetrized stars. It is also clear that $b \in \mathrm{St}(b)$ if and only if $\| b \| = 1$.

Let now X be a topological space and $B(X)$ the Banach space described in the introduction.

LEMMA 3.2. *Let X be countably compact*[10] *and let $f_1, f_2 \in B(X)$, $\| f_1 \| \leqq 1$, $\| f_2 \| \leqq 1$. We have $f_1 \in \mathrm{St}(f_2)$ if and only if there is an $x \in X$ such that*

$$f_1(x) = f_2(x) = \pm 1.$$

Proof clear.

LEMMA 3.3. *Let X be a topological space and let $f \in B(X)$. We have $|f| = 1$ if and only if $\mathrm{SSt}(f)$ is the whole surface of the unit sphere of $B(X)$.*

It is obvious that if $|f(x)| = 1$ for all $x \in X$ and $f_1 \in B(X)$ and $\| f_1 \| = 1$ then either $\| f + f_1 \| = 2$ or $\| f - f_1 \| = 2$ and therefore $f_1 \in \mathrm{SSt}(f)$.

Now suppose that $\inf |f(x)| = k < 1$. Define

$$f_1(x) = \frac{1 - |f(x)|}{1 - k}.$$

Clearly $f_1 \in B(X)$ and $\| f_1 \| = 1$. However

$$\| f + f_1 \| = \sup \left| \frac{f(x) - kf(x) + 1 - |f(x)|}{1 - k} \right| \leqq 1 + k < 2$$

because if $f(x) \geqq 0$ we have

$$|f(x) - kf(x) + 1 - |f(x)|| = |1 - kf(x)| \leqq 1 - k^2$$

and if $f(x) \leqq 0$ then

$$|f(x) - kf(x) + 1 - |f(x)|| = |1 - (2 - k)|f(x)|| \leqq (1 - k)^2.$$

Similarly $\| f - f_1 \| < 2$ and f_1 *non* $\in \mathrm{SSt}(f)$.

Lemma 3.2 shows that the elements of $B(X)$ such that $|f| = 1$ can be characterized intrinsically in $B(X)$. Hence the number of components of X can be determined by the properties of $B(X)$, since X will consist of n components if and only if there will be exactly 2^n continuous functions f such that $|f| = 1$.

4. The sets $M(f)$

Let $f \in B(X)$ and $\| f \| \leqq 1$. Define

$$M(f) = \{x \mid |f(x)| = 1\}.$$

[10] A space X is called countably compact if it is a Hausdorff space and if every countable covering of X by open sets contains a finite subcovering.

Clearly $M(f)$ is a closed subset of X. If $\|f\| < 1$ then obviously $M(f)$ is empty; if $\|f\| = 1$, $M(f)$ still may be empty unless X is countably compact.

LEMMA 4.1. *The sets $M(f)$ form a sublattice of the lattice of the closed subsets of X.*

In fact, we have $M(f_1) \cap M(f_2) = M(f)$ where $f(x) = \min [|f_1(x)|, |f_2(x)|]$ and $M(f_1) \cup M(f_2) = M(f)$ where $f(x) = \max [|f_1(x)|, |f_2(x)|]$.

LEMMA 4.2. *The space X is completely regular if and only if the sets $M(f)$ form a multiplicative base for the closed sets in X.*

If X is completely regular, then given a closed set $F \subset X$ and a point $x_0 \in X - F$, there is an $f \in B(X)$ such that $\|f\| \leqq 1$, $f(x_0) = 0$ and $f(x) = 1$ for $x \in F$. Hence $F \subset M(f)$ and x_0 *non* $\in M(f)$. Conversely, if $M(f)$ form a multiplicative base, then for a given closed set $F \subset X$ and a point $x_0 \in X - F$ there is an $f \in B(X)$, such that $\|f\| \leqq 1$, $F \subset M(f)$ and x_0 *non* $\in M(f)$. It follows that $|f(x_0)| < 1$ and $|f(x)| = 1$ for all $x \in F$; hence X is completely regular.

LEMMA 4.3.[11] *If X is normal then the sets $M(f)$ coincide with the closed G_δ subsets of X.*

LEMMA 4.4. *If X is countably compact then the relations*

$$M(f_1) \cap M(f_2) \neq 0 \quad \textit{and} \quad f_2 \in \mathrm{SSt}(f_1)$$

are equivalent for any $f_1, f_2 \in B(X)$ such that $\|f_1\| \leqq 1$, $\|f_2\| \leqq 1$.

In fact, $f_2 \in \mathrm{SSt}(f_1)$ means that $\|f_1 + f_2\| = 2$ or that $\|f_1 - f_2\| = 2$. Since X is countably compact this means that for some $x_0 \in X$ either $|f_1(x_0) + f_2(x_0)| = 2$ or $|f_1(x_0) - f_2(x_0)| = 2$. This is equivalent with $|f_1(x_0)| = |f_2(x_0)| = 1$ which means $M(f_1) \cap M(f_2) \neq 0$.

LEMMA 4.5. *If X is completely regular and countably compact then the relations*

$$M(f_1) \subset M(f_2) \quad \textit{and} \quad \mathrm{SSt}(f_1) \subset \mathrm{SSt}(f_2)$$

are equivalent for any $f_1, f_2 \in B(X)$ such that $\|f_1\| \leqq 1$, $\|f_2\| \leqq 1$.

Assume that $M(f_1) \subset M(f_2)$ and let $f \in \mathrm{SSt}(f_1)$. In view of Lemma 4.4 we have then $M(f) \cap M(f_1) \neq 0$; consequently $M(f) \cap M(f_2) \neq 0$ and byLemma 4.4 we have $f \in \mathrm{SSt}(f_2)$. Hence $\mathrm{SSt}(f_1) \subset \mathrm{SSt}(f_2)$.

Let now $x_0 \in M(f_1) - M(f_2)$. Since X is completely regular there is an $f \in B(X)$ such that $\|f\| \leqq 1$, $f(x_0) = 1$ and $f(x) = 0$ for all $x \in M(f_2)$. It follows that $x_0 \in M(f)$, therefore $M(f) \cap M(f_1) \neq 0$ and $f \in \mathrm{SSt}(f_1)$. Since $f(x) = 0$ for all $x \in M(f_2)$ we have $|f(x) \pm f_2(x)| < 2$ and since X is countably compact this implies that $\|f \pm f_2\| < 2$. Consequently $f \in \mathrm{SSt}(f_1) - \mathrm{SSt}(f_2)$.

From Lemmas 4.3 and 4.5 we obtain the following

THEOREM 4.6. *If X is normal and countably compact, then the sets $\mathrm{SSt}(f)$ ($f \in B(X)$, $\|f\| \leqq 1$) ordered by inclusion form a lattice isomorphic with the lattice of the closed G_δ subsets of X.*

5. The sets $Q(x)$

Given a point $x \in X$ we define

$$Q(x) = \{f \mid f \in B(X), \|f\| = 1, |f(x)| = 1\}.$$

[11] For the proof see N. Vedenisoff, *Fund. Math.*, 27 (1936), pp. 234–238.

LEMMA 5.1. *Let X be completely regular. Given $x \in X$ and $f \in B(X)$ such that $\|f\| \leqq 1$ we have $Q(x) \subset \mathrm{SSt}(f)$ if and only if $|f(x)| = 1$.*

If $|f(x)| = 1$ it follows directly from the definition that $Q(x) \subset \mathrm{SSt}(f)$. If now $|f(x)| < 1$ then there is an open set G such that $x \in G$ and $|f(x')| < 1 - \epsilon < 1$ for all $x' \in G$. Hence there is an $f_1 \in B(X)$ such that $\|f_1\| = 1$, $f_1(x) = 1$, and $f_1(x') = 0$ for all $x' \in X - G$. Consequently $f_1 \in Q(x)$ but since $\|f \pm f_1\| \leqq 2 - \epsilon$ we have f_1 *non* $\in \mathrm{SSt}(f)$.

THEOREM 5.2. *If X is compact, then a subset A of the unit sphere of $B(X)$ is of the form $Q(x)$ for some $x \in X$ if and only if $A = SA_1$ where A_1 is a maximal convex subset of the surface of the unit sphere in $B(X)$.*

This is an immediate consequence of the following

THEOREM 5.3. *If X is compact then a subset A of the surface of the unit sphere of $B(X)$ is maximal convex if and only if there is an $x_0 \in X$ and $\epsilon = \pm 1$ such that $A = \{f \mid f \in B(X), \|f\| = 1, f(x_0) = \epsilon\}$.*

We first show that if such x_0 and ϵ exist then A is maximal convex. It is obvious that A is convex. Let then A^* be a convex set contained in the unit sphere of $B(X)$, such that $A \subset A^*$ and $f \in A^* - A$. Since $f(x_0) \neq \epsilon$ and X is compact and therefore normal there is an $f_1 \in B(X)$ such that $\|f_1\| = 1$, $f_1(x_0) = \epsilon$ and $f_1(x) = 0$ for $x \in M_\epsilon(f)$. It follows that $|f(x) + f_1(x)| < 2$ for all x and therefore $\frac{1}{2}(f_1 + f)$ *non* $\in A^*$. However, $f_1 \in A^*$ and $f \in A^*$, hence A^* was not convex.

In order to prove that every maximal convex subset A of the unit sphere is of the form described in Theorem 5.3, it is enough to find an $x_0 \in X$ and an $\epsilon = \pm 1$ such that $f(x_0) = \epsilon$ for all f in A. If we denote

$$M_+(f) = \{x \mid f(x) = -1\} \qquad M_-(f) = \{x \mid f(x) = -1\}$$

this reduces to prove that either

$$\bigcap_{f \in A} M_+(f) \neq 0 \quad \text{or} \quad \bigcap_{f \in A} M_-(f) \neq 0.$$

Assume the contrary. Since the sets $M_+(f)$ and $M_-(f)$ are closed and X is compact, we would have then two finite sequences $f_1, f_2, \cdots, f_m, g_1, g_2, \cdots, g_n$ in A such that

$$\bigcap M_+(f_i) = 0 \quad \text{and} \quad \bigcap M_-(g_i) = 0.$$

Since A is convex we have $(f_1 + \cdots + f_m + g_1 + \cdots + g_n)/(n + m) \in A$, hence $\|f_1 + \cdots + f_m + g_1 + \cdots + g_n\| = n + m$, therefore there is an $x_1 \in X$ such that $f_1(x_1) = \cdots = f_m(x_1) = g_1(x_1) = \cdots = g_n(x_1) = \pm 1$ which leads to a contradiction.

6. Reconstruction of X from $B(X)$

Let B be a Banach space. A subset of B will be called an *ideal point*, and denoted by a Greek letter ξ if $\xi = SA$ where A is a maximal convex subset of the surface of the unit sphere of B. The totality of all the ideal points will be

noted by Ξ. A subset Φ of Ξ will be called an M-set if there is an $f \in B$ such that $\|f\| \leqq 1$ and

$$\Phi = \{\xi \mid \xi \subset \mathrm{SSt}(f)\}.$$

A subset of Ξ which is an intersection of M-sets will be called *closed*.

It is to be noted that with this definition of closed sets Ξ does not necessarily become a topological space. In fact, if Ξ were a topological space it would have to be a closed set and therefore an M-set. This would imply the existence of an element $f_1 \in B$ such that $\|f_1\| = 1$ and that $\Xi = \{\xi \mid \xi \subset \mathrm{SSt}(f_1)\}$. From this it follows easily that $\mathrm{SSt}(f_1)$ is the whole surface of the unit sphere of B. Of course, in most Banach spaces no such element exists.

THEOREM 6.1. *If X is a compact space then Ξ defined from the Banach space $B(X)$ is a topological space homeomorphic to X.*

Given $x \in X$ it follows from Theorem 5.2 that $Q(x)$ is an ideal point and that all the points of Ξ are obtained in this fashion. Hence taking

$$\xi(x) = Q(x)$$

we obtain a mapping of X onto Ξ.

We prove that ξ is $1 - 1$. Let $x_1, x_2 \in X$ and $x_1 \neq x_2$. Since X is completely regular there is an $f \in B(X)$ such that $\|f\| \leqq 1, f(x_1) = 1$ and $f(x_2) = 0$. Hence $f \in Q(x_1)$ but f *non* $\in Q(x_2)$, therefore $Q(x_1) \neq Q(x_2)$.

We now prove that ξ is a homeomorphism. Since the sets $M(f)$ form a multiplicative base for the closed sets in X (see Lemma 4.2) and by definition the M-sets form a similar base for the closed sets in Ξ, we shall prove that h is a homeomorphism by showing that if

$$\xi(F) = \Phi$$

then F is a set $M(f)$ if and only if Φ is an M-set.

Suppose that $F = M(f)$. It follows from Lemma 5.1 that $Q(x) \subset \mathrm{SSt}(f)$ if and only if $|f(x)| = 1$, i.e. if $x \in F$. Consequently

$$\Phi = \{\xi \mid \xi \subset \mathrm{SSt}(f)\}$$

and Φ is an M-set.

If f will vary over all of the unit sphere of $B(X)$, F will vary over all $M(f)$ sets and $\xi(F) = \Phi$ will vary over all M-set in Ξ. Since ξ is $1 - 1$, the proof is complete.

From Theorem 6.1 and from §2, we obtain

THEOREM 6.2. *If X is a completely regular space, then Ξ defined from the Banach space $B(X)$ is a topological space homeomorphic with the compacting $\beta(X)$ of X.*

If, in addition, X satisfies the first countability axiom, then X is homeomorphic with the subset of Ξ consisting of points which have a countable base.

7. Isometric mappings of $B(X)$ onto $B(Y)$

THEOREM 7.1. *Given two topological spaces X and Y, every isometric mapping T of $B(X)$ onto $B(Y)$ is of the form*

$$(*) \qquad Tf = g_0 + g_1 \cdot (T'f)$$

where $g_0, g_1 \in B(Y)$, $|g_1| = 1$ and T' is an isometry of $B(X)$ onto $B(Y)$ that is an algebraic and lattice isomorphism.

Define $T_1 f = Tf - T0$. Clearly T_1 is an isometry and $T_1 0 = 0$; hence by a theorem of S. Mazur and S. Ulam[12] T_1 and its inverse are linear. Consequently we have $T_1 \mathrm{St}(f) = \mathrm{St}(T_1 f)$. Consider the function 1 in $B(X)$, since $\mathrm{SSt}(1)$ is the whole surface of the unit sphere it follows that $\mathrm{SSt}(T_1 1)$ is the whole surface of the unit sphere in $B(Y)$ and by Lemma 3.3 $|T_1 1| = 1$. If we define $g_0 = T0$, $g_1 = T_1 1$ and $T'f = T_1 f / T_1 1$ we have (*) and T' is a linear isometry such that $T'0 = 0$ and $T'1 = 1$. All that remains to be proved is that T' preserves the lattice properties.

Suppose $f \geqq 0$, then[13]

$$T'f = T'\,||f|| + T'f - T'\,||f|| = ||f|| - T'(||f|| - f)$$
$$\geqq ||f|| - ||\,T'(||f|| - f)\,|| = ||f|| - ||\,||f|| - f\,|| \geqq 0$$

and $T'f \geqq 0$. The same holds for the inverse of T' and therefore T' is a lattice isomorphism.

THEOREM 7.2. *Let X and Y be two compact spaces. Every isometric mapping T of $B(X)$ onto $B(Y)$ is of the form*

$$Tf = g_0 + g_1 \cdot (fh)$$

where $g_0, g_1 \in B(Y)$, $|g_1| = 1$ and h is a homeomorphism $h(Y) = X$.

If $T0 = 0$ then $g_0 = 0$ and

$$Tf = g_1 \cdot (fh).$$

If $T0 = 0$ and $T1 = 1$, then $g_0 = 0$, $g_1 = 1$ and

$$Tf = fh.$$

The same holds if X and Y are completely regular and satisfy the first countability axiom; in particular, if X and Y are metric spaces.

In view of the previous theorem we may restrict ourselves to the case when $T0 = 0$ and $T1 = 1$. The isometry T is then linear and preserves all the lattice properties. In particular

$$(**) \qquad T|f| = |Tf|.$$

We first discuss the case when X and Y are compact. Let Ξ_X and Ξ_Y be the corresponding ideal spaces and ξ_X and ξ_Y the homeomorphisms $\xi_X(X) = \Xi_X$

[12] C. R. Paris 194 (1932), pp. 946–948; also Banach *loc. cit.* p. 168.

[13] A real number a is identified here with the function of $f \in B(x)$ such that $f(x) = a$ for all $x \in X$.

and $\xi_Y(Y) = \Xi_Y$. Since Ξ_X and Ξ_Y were defined intrinsically using $B(X)$ and $B(Y)$ it is clear that T induces a homeomorphism between Ξ_X and Ξ_Y which we will also denote by T; $T(\Xi_X) = \Xi_Y$. Define

$$h(y) = \xi_X^{-1} T^{-1} \xi_Y(y) \qquad \text{for } y \in Y.$$

Clearly h is a homeomorphism $h(Y) = X$. It remains to prove that $Tf = fh$ for all $f \in B(X)$.

We shall first prove that

$$(^{**}_{*}) \qquad \textit{if} \quad \|f\| \leqq 1 \quad \textit{then} \quad M(f) = h[M(Tf)].$$

This follows immediately from the remark that $\xi_X[M(f)]$ and $\xi_Y[M(Tf)]$ are the M-sets

$$\xi_X[M(f)] = \{\xi \mid \xi \subset \mathrm{SSt}(f)\}$$

$$\xi_Y[M(Tf)] = \{\xi \mid \xi \subset \mathrm{SSt}(Tf)\} = \{\xi \mid \xi \subset T\mathrm{SSt}(f)\},$$

hence $T\xi_X[M(f)] = \xi_Y[M(Tf)]$ which implies $(^{**}_{*})$ in view of the definition of h.

Now let $f \in B(X)$, $x \in X$, $h(y) = x$ and $f(x) = a$. Consider

$$f_1 = 1 - \frac{|f - a|}{\|f - a\|}.$$

Because of (**) we have

$$Tf_1 = 1 - \frac{|Tf - a|}{\|Tf - a\|}.$$

Since $x \in M(f_1)$ we shall have in view of $(^{**}_{*})$ $y \in M(Tf_1)$, hence $|(Tf_1)(y)| = 1$ which implies $(Tf)(y) = a$. This shows that $Tf = hf$ which proves the theorem in the compact case.

If now X and Y are completely regular, we consider the compactings $\beta(X)$ and $\beta(Y)$. Since $B[\beta(X)] = B(X)$ and $B[\beta(Y)] = B(Y)$ we have an isometry T mapping $B[\beta(X)]$ onto $B[\beta(Y)]$ and hence a homeomorphism $h[\beta(Y)] = \beta(X)$ such that $Tf = fh$. If X and Y satisfy the first countability axiom, then they are the sets of points at which the first countability axiom holds in $\beta(X)$ and $\beta(Y)$ respectively. Therefore $h(Y) = X$.

8. Discussion of Banach's proof

Theorem 7.2 was first established by Banach[2] for X and Y compact metric. His proof is based on the concept of a peak function. A function $f \in B(X)$ is called a *peak function* with the point $x_0 \in X$ as *peak* if $|f(x)| < |f(x_0)|$ for all $x \in X$, $x \neq x_0$. Banach proves it if X is compact metric, then $f \in B(X)$ is a peak function if and only if $f \neq 0$ and the limit

$$\lim_{t \to 0} \frac{\|f + tf_1\| - \|f\|}{t}$$

exists for every $f_1 \in B(X)$. Having such an intrinsic characterization of peak functions, it is easy to see that an isometry $TB(X) = B(Y)$ such that $T0 = 0$ must carry peak function into peak functions and this way a homeomorphism between X and Y can readily be established.

We propose to replace Banach's characterization of peak functions by the following one which seems to be more geometrical and natural:

THEOREM 8.1. *Let X be completely regular and countably compact. A function $f \in B(X)$ such that $\|f\| = 1$ is a peak function if and only if* $\mathrm{St}(f)$ *is convex.*

If f is a peak function then there is an $x_0 \in X$ such that

$$|f(x_0)| = 1, \qquad |f(x)| < 1 \qquad \text{for } x \neq x_0.$$

Since X is countably compact it follows from Lemma 3.2 that

$$\mathrm{St}(f) = \{f_1 \mid f_1 \in B(X), \|f_1\| = 1, f_1(x_0) = f(x_0)\}$$

and this set is clearly convex.

If f is not a peak function, then since X is countably compact f assumes its maximum and therefore there are at least two distinct points $x_1, x_2 \in X$ such that

$$|f(x_1)| = |f(x_2)| = 1.$$

Since X is completely regular there are two open sets G_1 and G_2 such that

$$x_1 \in G_1, \qquad x_2 \in G_2 \quad \text{and} \quad G_1 \cap G_2 = 0$$

and two functions $f_1, f_2 \in B(X)$ such that $\|f_1\| = \|f_2\| = 1$ and

$$f_i(x_i) = f(x_i), \qquad f_i(x) = 0 \qquad \text{for } x \in X - G_i,\ i = 1, 2.$$

Clearly $\|f + f_1\| = 2$, $\|f + f_2\| = 2$ and $\|f_1 + f_2\| = 1$ since $X = (X - G_1) \cup (X - G_2)$. Hence $f_1 \in \mathrm{St}(f)$, $f_2 \in \mathrm{St}(f)$ and $\frac{1}{2}(f_1 + f_2)$ *non* $\in \mathrm{St}(f)$. Hence $\mathrm{St}(f)$ is not convex.

Having an intrinsic characterization of the peak functions we can carry out Banach's proof provided we know that for every $x_0 \in X$ there is a peak function with x_0 as peak. It can be easily seen that this is true if and only if X is completely regular and every point of X is a G_δ. If in addition X is countably compact, this last condition becomes equivalent with the first countability axiom. Hence we see that Banach's proof can be carried out for completely regular, countably compact spaces satisfying the first countability axiom.

9. Representations of $B(X)$ as a direct sum

Given two linear subspaces B_1, B_2 of a Banach space B we say that B is the direct sum

$$B = B_1 + B_2$$

if every $b \in B$ admits a unique decomposition

$$b = b_1 + b_2, \qquad b_1 \in B, \qquad b_2 \in B_2$$

and if[14]

$$\| b \| = \max (\| b_1 \|, \| b_2 \|).$$

It is easy to prove that if $B = B_1 + B_2$ then B_1 and B_2 are closed subspaces of B.

Let X be a topological space and let

$$(*) \qquad X_1 = X_1 \cup X_2, \qquad X_1 \cap X_2 = 0, \qquad \bar{X}_1 = X_1, \qquad \bar{X}_2 = X_2.$$

Let $B(X_1)$, $B(X_2)$ be the subspaces of $B(X)$ defined as follows

$$B(X_i) = \{f \mid f(x) = 0 \text{ for all } x \text{ non } \epsilon\, X_i\}, \qquad i = 1, 2.$$

We verify easily that $B(X) = B(X_1) + B(X_2)$.

LEMMA 9.1. *Every decomposition $X = X_1 \cup X_2$ of X into two disjoint closed sets determines a direct sum decomposition $B(X) = B(X_1) + B(X_2)$.*

We shall show that the converse is also true.

THEOREM 9.2. *Let X be a topological space. For every decomposition*

$$B(X) = B_1 + B_2$$

as a direct sum there is a decomposition

$$X = X_1 \cup X_2$$

into disjoint closed sets such that $B_1 = B(X_1)$ and $B_2 = B(X_2)$.

We first assume that X is compact. Since $B(X) = B_1 + B_2$ every $f \epsilon B(X)$ has a unique decomposition

$$f = f_1 + f_2, \qquad f_1 \epsilon B_1, \qquad f_2 \epsilon B_2.$$

Define

$$Tf = f_1 - f_2.$$

Clearly T is linear and

$$\| Tf \| = \| f \|,$$

hence $TB(X) = B(X)$ is an isometry such that $T0 = 0$ and because of Theorem 7.2 we have

$$Tf = g_1 \cdot (fh)$$

where $g_1 \epsilon B(X)$, $| g_1 | = 1$, and h is a homeomorphic mapping of X onto itself.

We shall prove that $h(x) = x$ for all $x \epsilon X$. Assume to the contrary that $h(x_0) = x_1 \neq x_0$ for some $x_0 \epsilon X$. There is then an open set $G \subset X$ such that

$$x_0 \epsilon G, \qquad G \cap h(G) = 0.$$

---

[14] The definition of the direct sum essentially depends upon the expression chosen for the norm. The most commonly used are $\| b \| = (\| b_1 \|^p + \| b_2 \|^p)^{1/p}$ for $p \geqq 1$. The expression in the text is obtained by letting $p \to \infty$.

Since X is compact it is also completely regular and hence there is an $f \in B(X)$ such that

$$f(x_0) = 1, \qquad f(x) = 0 \text{ for all } x \in X - G, \qquad \|f\| = 1.$$

Since $f = f_1 + f_2$ where $f_1 \in B_1$, $f_2 \in B_2$ and $\|f\| = \max(\|f_1\|, \|f_2\|)$ we may assume that $\|f_1\| = 1$. Since X is compact there is an $x' \in X$ such that

$$|f_1(x')| = 1.$$

Since

$$|f_1(x') + f_2(x')| = |f(x')| \leqq \|f\| = 1$$

$$|f_1(x') - f_2(x')| = |Tf(x')| \leqq \|Tf\| = 1$$

we must have $f_2(x') = 0$. Consequently $f(x') = f_1(x') = \pm 1$ and therefore $x' \in G$. Since $f_1 \in B_1$ we have $Tf_1 = f_1$ and therefore

$$f_1 = g_1 \cdot (f_1 h).$$

Consequently $|f_1(h(x'))| = |f_1(x')| = 1$ and by the previous argument $h(x') \in G$. Hence $G \cap h(G) \neq 0$, a contradiction.

Having proved that $h(x) = x$ for all $x \in X$ we have

$$Tf = g_1 \cdot f \quad \text{where} \quad |g_1| = 1.$$

Define

$$X_1 = \{x \mid g_1(x) = 1\}, \qquad X_2 = \{x \mid g_1(x) = -1\}.$$

Clearly X_1, X_2 are closed and $X = X_1 \cup X_2$. Given $f \in B(X)$ we have $f \in B_1$ if and only if $Tf = f$, this means that $f = g_1 \cdot f$ and this holds if and only if $f(x) = 0$ whenever $g_1(x) = -1$. Hence we see that $B_1 = B(X_1)$. Analogously $B_2 = B(X_2)$.

Next we assume that X is completely regular. Let $\beta(X)$ be the compacting of X. Since $B(\beta(X)) = B(X)$ we have a direct sum decomposition $B(\beta(X)) = B_1 + B_2$ and hence there is a decomposition

$$\beta(X) = Y_1 \cup Y_2$$

into disjoint closed sets such that $B_1 = B(Y_1)$ and $B_2 = B(Y_2)$. Let

$$X_1 = X \cap Y_1, \qquad X_2 = X \cap Y_2.$$

From the definition of the compacting it is clear that $Y_1 = \beta(X_1)$ and $Y_2 = \beta(X_2)$. Hence $B_1 = B(X_1)$, $B_2 = B(X_2)$.

If now X is an arbitrary topological space we consider the completely regular space $\rho(X)$ defined by Čech.[8] Again $B(\rho(X)) = B(X)$ and hence we have the direct sum decomposition $B(\rho(X)) = B_1 + B_2$. Since the theorem is proved for completely regular spaces there is a decomposition

$$\rho(X) = Y_1 \cup Y_2$$

into disjoint closed sets such that $B_1 = B(Y_1)$ and $B_2 = B(Y_2)$. From the definition of $\rho(X)$ it is clear that there is a decomposition

$$X = X_1 \cup X_2$$

into disjoint closed sets such that $Y_1 = \rho(X_1)$ and $Y_2 = \rho(X_2)$. Consequently $B_1 = B(X_1)$ and $B_2 = B(X_2)$.

From Theorem 9.2 we obtain the following corollaries:

COROLLARY 9.3. *X is connected if and only if $B(X)$ is indecomposable as a direct sum.*

COROLLARY 9.4. *X contains at least n components if and only if $B(X)$ is a direct sum of n summands.*

COROLLARY 9.5. *X contains an isolated point if and only if $B(X)$ admits the straight line as a direct summand.*

UNIVERSITY OF MICHIGAN.

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

SIMPLIZIALZERLEGUNGEN VON BESCHRÄNKTER FLACHHEIT[1]

BY HANS FREUDENTHAL

(Received July 25, 1941)

Wir beantworten eine Frage von Herrn Prof. L. E. J. Brouwer nach einer einfachen Konstruktion einer unendlichen Folge beliebig feiner Simplizialzerlegungen eines Polytops, deren jede eine Unterteilung der vorangehenden ist, und bei der die auftretenden Teilsimplexe nicht beliebig flach werden dürfen, d. h. bei der der Quotient

$$c^r/v$$

(c = Durchmesser = größte Kantenlänge des Simplexes, v = Volumen des Simplexes) für alle r-dimensionalen Simplexe gleichmäßig beschränkt bleibt. Derartige Zerlegungsfolgen braucht man in der Analysis und auch im Grenzgebiet zwischen Analysis und Topologie.

Die Konstruktion, die wir hier angeben wollen, verläuft ähnlich wie unsere Simplizialzerlegung des Cartesischen Produkts zweier Simplexe.[2]

1

Simplexe sollen hier immer mit einer *festen* Reihenfolge der Ecken gegeben sein, Parallelotope immer unter Auszeichnung einer Ecke und einer Anordnung der von dieser Ecke ausgehenden Kanten.

Das Simplex T habe die Ecken

$$e_0, \cdots, e_r.$$

Wir können T auch beschreiben durch die Vektoren

$$x_0 = e_0, \qquad x_i = e_i - e_{i-1} \qquad (i = 1, \cdots, r).$$

Sei π eine Permutation $p_1, \cdots, p_r$ der Zahlen $1, \cdots, r$. Die Punkte

$$e_i = x_0 + \sum_{\nu=1}^{r} x_{p_\nu} \qquad (i = 0, \cdots, r)$$

(als Ecken) erzeugen ein Simplex T^π. Ist π_0 die identische Permutation der Zahlen $1, \cdots, r$, so ist $T^{\pi_0} = T$. Die $r!$ Simplexe T^π heißen *untereinander konjugiert*.

T^π ist die Gesamtheit der Punkte

$$T^\pi \cdots \sum_{\nu=1}^{r} \lambda_\nu e_{p_\nu} \quad \text{mit} \quad \lambda_\nu \geqq 0, \qquad \sum_{\nu=1}^{r} \lambda_\nu = 1,$$

---

[1] Diese Note stimmt im Wesentlichen überein mit einer im März 1939 bei den Fundamenta Mathematicae eingereichten Note die dort nicht mehr erschienen ist. Das Ergebnis habe ich bereits anderswo verwendet [Proceedings Amsterdam 42 (1939), 880–901].

[2] Fund. Math. 29 (1937), 138–144.

oder, was auf dasselbe hinauskommt,

$$T^\pi \cdots x_0 + \sum_{\nu=1}^{r} \alpha_\nu x_{p_\nu} \quad \text{mit} \quad 1 \geqq \alpha_1 \geqq \cdots \alpha_r \geqq 0.$$

Daraus folgt: *Die T^π bilden eine simpliziale Teilung des Parallelotops P,*

$$P \cdots x_0 + \sum \alpha_\nu x_\nu , \qquad 0 \leqq \alpha_\nu \leqq 1.$$

Umgekehrt wird durch den Parallelotop P, von dem die Ecke x_0 und die Kantenvektoren $x_1 , \cdots , x_r$ gegeben sind, und durch die Permutation π eindeutig das Simplex T^π bestimmt.

2

Aus dem Parallelotop P entstehen durch Halbierung quer zu allen Kanten die Parallelotope

$$P_\sigma \cdots x_{\sigma,0} + \tfrac{1}{2} \sum \alpha_\nu x_\nu ;$$

hier sei σ ein System

$$s_1 , \cdots , s_r$$

von Zahlen s_ν ,

$$s_\nu = 0 \text{ oder } 1,$$

und

$$x_{\sigma,0} = x_0 + \tfrac{1}{2} \sum s_\nu x_\nu .$$

In derselben Weise wie P in die Simplexe T^π ist jedes P_σ zerlegt in die Simplexe $(T_\sigma)^\pi$.—*T^π und $(T_\sigma)^\pi$ sind homothetisch* (1:2).

Sei π wieder die Permutation $p_1 , \cdots , p_r$ der $1, \cdots , r$ und sei $\pi^* = \pi^\sigma$ die Permutation $p_1^* , \cdots , p_r^*$, die entsteht, wenn man erst (in der durch π gegebenen Reihenfolge) alle p_ν mit $s_{p_\nu} = 1$ auszieht, und dann alle mit $s_{p_\nu} = 0$, also:

für $s_a > s_b$ kommt a vor b in π^*,

für $s_a = s_b$ und a vor b in π kommt auch a vor b in π^*.

Wir behaupten: *$(T_\sigma)^\pi$ liegt ganz in T^{π^σ}.*

In der Tat ist

$$(T_\sigma)^\pi \cdots x_0 + \tfrac{1}{2} \sum_{\nu=1}^{r} s_\nu x_\nu + \tfrac{1}{2} \sum_{\nu=1}^{r} \alpha_\nu x_{p_\nu}$$

$$= x_0 + \tfrac{1}{2} \sum{}' (\alpha_\nu + \tfrac{1}{2}) x_{p_\nu} + \tfrac{1}{2} \sum{}'' \alpha_\nu x_{p_\nu} \qquad (1 \geqq \alpha_1 \geqq \cdots \geqq \alpha_r \geqq 0)$$

(wo $\sum'$ zu erstrecken ist über alle ν mit $s_{p_\nu} = 1$ und $\sum''$ über alle ν mit $s_{p_\nu} = 0$)

$$= x_0 + \sum_{\nu=1}^{r} \beta_\nu x_{p_\nu^*} \qquad (1 \geqq \beta_1 \geqq \cdots \geqq \beta_u \geqq \tfrac{1}{2} \geqq \beta_{u+1} \geqq \cdots \geqq \beta_r \geqq 0)$$

(wo u die Zahl der Einsen unter den s_ν ist). Hieraus folgt unsere Behauptung.

Weiter folgt hieraus: *Die $(T_\sigma)^\pi$ mit $\pi^\sigma = \pi^*$ bilden eine Simplizialzerlegung $Z(T)$ von T.*

3

Sei $e_{ij}^{\pi^*}$ der Mittelpunkt der Strecke $e_i^{\pi^*} e_j^{\pi^*}$, $e_{ii}^{\pi^*} = e_i^{\pi^*}$,

$$e_{ij}^{\pi^*} = \frac{1}{2}\left(x_0 + \sum_{\nu=1}^{i} x_{p_\nu^\pi} + x_0 + \sum_{\nu=1}^{j} x_{p_\nu^\pi}\right)$$

$$= x_0 + \sum_{\nu=1}^{i} x_{p_\nu^\pi} + \frac{1}{2} \sum_{\nu=i+1}^{i} x_{p_\nu^\pi} \qquad (i \leqq j).$$

Seien unter den s_{p_ν} mit $\nu \leqq k$

k' Einsen und k'' Nullen.

Die k^{te} Ecke von $(T_\sigma)^\pi$ erhält man für $\alpha_1 = \cdots = \alpha_k = 1$, $\alpha_{k+1} = \cdots = \alpha_r = 0$; sie lautet also

$$x_0 + \sum_{\nu=1}^{k'} x_{p_\nu^\pi} + \frac{1}{2} \sum_{\nu=l'+1}^{l''+u} x_{p_\nu^\pi} = e_{ij}^{\pi^*} \qquad (i = k', j = k'' + u).$$

Daraus folgt für jedes Teilsimplex $(T_\sigma)^\pi$ von T^{π^σ}:

0^{te} *Ecke ist* $e_{0u}^{\pi^\sigma}$;

auf die Ecke $e_{ij}^{\pi^\sigma}$ *folgt die Ecke* $e_{i+1,j}^{\pi^\sigma}$ *oder die Ecke* $e_{i,j+1}^{\pi^\sigma}$;

der erste Index ist stets $\leqq u$, *der zweite* $\leqq r$.

Umgekehrt sind die Teilsimplexe $(T_\sigma)^\pi$ von T^{π^σ} durch diese Eigenschaften auch gekennzeichnet.

4

Wir achten nun nur noch auf $T = T^{\pi_0}$ und lassen den Index π_0 im Folgenden weg. Wir können dann die Teilung $Z(T)$ von T auch so beschreiben:

Die Ecken von $Z(T)$ *sind die* e_{ij} ;

dann und nur dann bilden e_{ij} *und* $e_{i'j'}$ *ein eindimensionales Simplex, wenn die Paare* ij *und* $i'j'$ *einander nicht trennen*;

eine Menge von e_{ij} *bildet dann und nur dann ein Simplex von* $Z(T)$, *wenn ihre Elemente zu je zweien Simplexe von* $Z(T)$ *bilden*;

in den Simplexen von $Z(T)$ *sind die Ecken nach steigenden* $i + j$ *angeordnet.*

Wir bemerken weiter:

Die Simplexe von $Z(T)$ *sind ähnlich der zu* T *konjugierten.*

Sei nun ein endlicher Polytop R gegeben! Wir legen eine Reihenfolge der Ecken zugrunde und bilden die Teilung $Z(R)$, indem wir auf jedes Simplex von R den Prozeß Z anwenden. Die Ecken von $Z(R)$ ordnen wir lexikographisch. Indem wir den Prozeß Z unbegrenzt wiederholen, erhalten wir eine Unterteilungsfolge. Alle dabei auftretenden Simplexe sind den (endlich vielen) konjugierten der Simplexe von R ähnlich, und da ähnliche Simplexe dieselbe Flachheit

$$c^r/v$$

besitzen, bleibt, wie wir es wünschten, die Flachheit gleichmäßig beschränkt.

AMSTERDAM

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

A SIMPLIFIED PROOF FOR THE RESOLUTION OF SINGULARITIES OF AN ALGEBRAIC SURFACE

BY OSCAR ZARISKI

(Received March 19, 1942)

1. Introduction

We presuppose the theorem of local uniformization, as proved elsewhere.[1] By this theorem, any valuation of a field Σ/K of algebraic functions of r independent variables, over a given ground field K of characteristic zero, can be "uniformized" on a suitable projective model V of Σ, i.e. the center of the valuation on a suitable model V will be a simple subvariety of V.[2] If the ground field K (the field of coefficients) is the field of complex numbers (the classical case), then the above result, in conjunction with the bicompactness of the Riemann manifold $\mathbf{M}$ of Σ/K,[3] implies the existence of a *finite* set of models of Σ, say

$$V_1, V_2, \cdots, V_n,$$

such that any zero-dimensional valuation of Σ/K is uniformized on at least one of the models V_i of the set.[4] Any finite set of projective models with the above property shall be called a *resolving system* of the Riemann manifold $\mathbf{M}$.

In the case of abstract varieties, where we cannot fall back on topology, it is necessary to give an algebraic proof of the existence of resolving systems of $\mathbf{M}$. In the special case of algebraic surfaces the algebraic proof of the existence of resolving systems is strikingly simple (see section 6 of this paper). The general case of algebraic varieties will be treated in a subsequent paper.

The theorem of the resolution of singularities of an algebraic variety can be formulated in terms of resolving systems, as follows: *There exists a resolving system of the Riemann manifold* $\mathbf{M}$ *which consists of only one model.* In view of the existence of resolving systems, this theorem will be established if it can be

[1] For the case of algebraic surfaces, see our paper "The reduction of singularities of an algebraic surface", Annals of Mathematics, vol. 40 (July, 1939), pp. 649–659. For the general case of varieties of any dimension, see our paper "Local uniformization on algebraic varieties", Annals of Mathematics, vol. 41 (October, 1940), pp. 852–896. These two papers will be referred to respectively as "Reduction" and "Uniformization".

[2] For the definition of the center of a valuation see "Uniformization", p. 857.

[3] $\mathbf{M}$ is the totality of all *zero-dimensional* valuations (or *places*) of Σ/K; see "Uniformization", p. 855.

[4] It follows then necessarily that *any* valuation (whether of dimension zero or greater than zero) will be uniformized on at least one of the models V_i. To see this, it is only necessary to observe that: a) If B is any valuation of dimension > 0, then there exist zero-dimensional valuations compounded with B; b) if B_0 is such a zero-dimensional valuation then on every model V of Σ the center of B_0 will lie on the center of B; c) if an irreducible subvariety W of V contains a simple point, then W itself is simple.

proved that *the existence of a resolving system of* **M** *consisting of n models* ($n > 1$) *implies the existence of a resolving system of* **M** *consisting of* $n - 1$ *models.* To carry out this induction from n to $n - 1$, it is sufficient to prove the *fundamental theorem* which we are now going to state.

Let N be an arbitrary subset of **M**, i.e. let N be an arbitrary set of places of our field Σ/K. In the same fashion as we have defined resolving system of **M**, we can define resolving systems of N. A resolving system of N will be therefore any finite set of models of Σ such that any valuation in N has a simple center on at least one of the models in the set. In particular, if a resolving system of N consists of only one model, this model shall be called a *resolving model for* N.

FUNDAMENTAL THEOREM. *If N is an arbitrary subset of* **M** *and if there exists a resolving system of N consisting of two models, then there also exists a resolving model for N.*

From the fundamental theorem, the theorem on the resolution of singularities follows immediately. For, let $V_1, \cdots, V_n$ be a resolving system of **M**. Let N be the subset of **M** consisting of those valuations which have a singular center on *each* of the models $V_1, \cdots, V_{n-2}$. It is clear that the pair V_{n-1}, V_n constitutes a resolving system for N. If we assume the fundamental theorem, then there exists a resolving model V'_{n-1} for N. The $n - 1$ models $V_1, \cdots, V_{n-2}, V'_{n-1}$ constitute a resolving system for **M**, and our induction from n to $n - 1$ is complete.

The aim of this paper is to give a proof of the fundamental theorem in the case of algebraic surfaces. Our present proof for the resolution of singularities is much simpler than our earlier proof (see "Reduction") and is also more general, since at present we do not assume that the ground field K is algebraically closed. In the course of the proof we have to make use of certain properties of fundamental loci of birational correspondences. This preliminary material, strictly confined to the precise needs of our proof, is presented in the next section. In a forthcoming paper dealing with the general theory of birational correspondences we study these properties systematically. A brief account of this general theory will be found in my address "Normal varieties and birational correspondences" (delivered before the annual meeting of the Society in Bethlehem in 1941) which has appeared in the Bulletin of the American Mathematical Society, Vol. 48, No. 6 (June 1942).

I. PRELIMINARY CONCEPTS

2. Birational correspondences

Let V and V' be two models of our field Σ/K.

DEFINITION. Two points, P, P' of V and V' respectively are corresponding points if there exists a valuation of Σ/K such that its center on V is the point P and its center on V' is the point P'.[5]

[5] Compare with p. 666 of "Reduction". Our study of the general theory of birational correspondences is based on this valuation-theoretic definition.

We say that P is a *fundamental point* of the birational correspondence between V and V', if there exists a corresponding point P' on V' such that $Q(P) \not\supseteq Q(P')$ (here $Q(P)$ denotes the quotient ring of P).

Suppose that there exists a point P' which corresponds to P and which is such that $Q(P') \subseteq Q(P)$. If v is a valuation of centers P and P' on V and V' respectively and if $\mathfrak{P}_v$ denotes the prime ideal of v, then $\mathfrak{P}_v \cap Q(P) = \mathfrak{p}$ and $\mathfrak{P}_v \cap Q(P') = \mathfrak{p}'$, where $\mathfrak{p}(\mathfrak{p}')$ is the ideal of non units in $Q(P)(Q(P'))$. Hence $\mathfrak{p} \cap Q(P') = \mathfrak{p}'$, and from this it follows that *any* valuation whose center on V is P has P' as center on V', i.e. *P' is the only point of V' which corresponds* to P. Therefore, if P is a point of V which is not fundamental, then to P there corresponds a unique point P' on V', and we have $Q(P') \subseteq Q(P)$. On the other hand, if P is fundamental, then it follows that $Q(P) \not\supseteq Q(P')$ for any point P' which corresponds to P.

In the case of surfaces we have proved elsewhere that if V and V' are *normal* surfaces and if to P there corresponds a finite number of points on V', then P is not fundamental.[6] Therefore, if P is fundamental, the locus of corresponding points P' is a variety of dimension 1.[7]

If neither V nor V' carry fundamental points, then the birational correspondence between V and V' shall be called *regular*.

In any case, the number of fundamental points on either surface is always finite.

Let $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$ be the homogeneous coördinates of the general point of V, and let $\eta_0^*, \eta_1^*, \cdots, \eta_m^*$ be the homogeneous coördinates of the general point[8] of V'. The $(n+1)(m+1)$ products

$$\omega_{ij}^* = \xi_i^* \eta_j^*$$

can be regarded as homogeneous coördinates of the general point of a variety V^* which is birationally equivalent to V and to V': for the quotient of any two ω^*'s, say $\omega_{ij}^*/\omega_{\alpha\beta}^*$, is equal to $\xi_i^*/\xi_\alpha^* \cdot \eta_j^*/\eta_\beta^*$, and is therefore an element of the field Σ. Moreover, the non-homogeneous coördinates of the general point of V^*, i.e. the quotients of the ω^*'s by a fixed ω^*, say by ω_{00}^*, generate the field Σ/K, since the quotients ξ_i^*/ξ_0^* (and also the quotients η_j^*/η_0^*) are among them. The variety V^* is called the *join* of V and V'. It is clear that the ring

---

[6] See "Reduction", p. 688, Theorem 5, but the proof is much simpler and is as follows. If $P_1', \cdots, P_h'$ are the points on V' which correspond to P, then for any valuation v whose center is P it must be true that the valuation ring of v contains one of h quotient rings $Q(P_i')$, whence it also contains the intersection $\mathfrak{J}$ of these quotient rings. Since $Q(P)$ is integrally closed, it is the intersection of the above mentioned valuation rings. Hence $Q(P) \supseteq \mathfrak{J}$. If $\mathfrak{p} \cap \mathfrak{J} = \bar{\mathfrak{p}}$, then $\bar{\mathfrak{p}}$ is one of the h prime maximal ideals of $\mathfrak{J}$, say the one relative to the point P_1', and it is clear that any valuation of center P will have center P_1' on V', i.e. $h = 1$. [For the definition of normal varieties, see our paper "Some results in the arithmetic theory of algebraic varieties", p. 282, American Journal of Mathematics, vol. 61 (April 1939). This paper will be referred to as "Results".]

[7] See "Reduction", p. 667.

[8] For the concept of homogeneous coördinates of the general point of an irreducible algebraic variety, see "Results" p. 284.

$\mathsf{K}[\omega_{10}^*/\omega_{00}^*, \cdots, \omega_{nm}^*/\omega_{00}^*]$ of polynomials in these non-homogeneous coördinates —briefly: the ring of non-homogeneous coördinates—coincides with the ring $\mathsf{K}[\xi_1^*/\xi_0^*, \cdots, \xi_n^*/\xi_0^*, \eta_1^*/\eta_0^*, \cdots, \eta_m^*/\eta_0^*]$ and is therefore the join of the corresponding rings of non-homogeneous coördinates of the general points of V and V'. From this one concludes immediately that the birational correspondence between V^* and either one of the two given models V, V' has no fundamental points on V^*.

It follows that each point P^* of V^* represents a unique pair (P, P') of *corresponding* points of V and V'. Conversely, each such pair (P, P') is represented by at least one point P^* of the join V^*. For this reason the join V^* is often referred to in the literature as the variety of pairs of corresponding points of V and V'. However, this last term may be misleading when we deal with a ground field K which is not algebraically closed. For in this case (and only in this case) it may happen that one and the same pair (P, P') of corresponding points of V and V' is represented by (or, we may say, splits into) more than one point of V^*.

The join of *any finite number* of models $V_1, \cdots, V_h$ can be defined as the join of V_h and of the join of $V_1, \cdots, V_{h-1}$. It is seen immediately that this definition is symmetric in $V_1, \cdots, V_h$.

3. Quadratic transformations

Let P be a point of V and let $\mathfrak{p}^*$ be the corresponding prime homogeneous ideal (of projective dimension zero) in the ring $\mathsf{K}[\xi_0^*, \xi_1^*, \cdots, \xi_n^*]$. Let $\eta_0^*, \eta_1^*, \cdots, \eta_m^*$ be a set of forms in $\xi_0^*, \xi_1^*, \cdots, \xi_n^*$, *of like degree*, such that the ideal $(\eta_0^*, \eta_1^*, \cdots, \eta_m^*)$ differs from $\mathfrak{p}^*$ only by an irrelevant primary component i.e. a component belonging to the irrelevant prime ideal $\mathfrak{p}_0 = (\xi_0^*, \xi_1^*, \cdots, \xi_n^*)$.[9] By a *quadratic transformation of center P* we mean the birational transformation which carries V into the variety $\bar{V}$ whose general point has the following homogeneous coördinates:

$$\omega_{ij}^* = \xi_i^* \eta_j^*, \qquad i = 0, 1, \cdots, n; j = 0, 1, \cdots, m. \tag{1}$$

This transformation depends of course on the choice of the forms η_j^*, but in a non-essential fashion. Namely, if $\bar{V}_1$ is the transform of V by a quadratic transformation, relative to another set of forms, *then $\bar{V}$ and $\bar{V}_1$ are in regular birational correspondence.*

PROOF. Let $\zeta_0^*, \zeta_1^*, \cdots, \zeta_\mu^*$ be the set of forms which defines $\bar{V}_1$, and let $\bar{\omega}_{ij}^* = \xi_i^* \zeta_j^*$. Let: α = degree η_j^*, β = degree ζ_j^*, $\mathfrak{A} = (\eta_0^*, \cdots, \eta_m^*)$, $\mathfrak{B} = (\zeta_0^*, \cdots, \zeta_\mu^*)$. Since the ideals $\mathfrak{A}$ and $\mathfrak{B}$ differ only in their primary irrelevant components, we will have for sufficiently high integers a and b: $\mathfrak{A} \cdot \mathfrak{p}_0^a \equiv 0(\mathfrak{B})$, $\mathfrak{B} \cdot \mathfrak{p}_0^b \equiv 0(\mathfrak{A})$. Select a and b so as to have $\alpha + a = \beta + b$.

[9] Let $\mathfrak{p}_0^* = (\phi_0, \phi_1, \cdots, \phi_s)$, where the ϕ_i are forms in the ξ^*'s and let ν_i = degree of ϕ_i, ν = max. $(\nu_0, \nu_1, \cdots, \nu_s)$. Then the forms $\xi_j^{*\nu-\nu_i}\phi_i$, $j = 0, 1, \cdots, n$, $i = 0, 1, \cdots, s$, satisfy the desired condition.

The elements ω_{ij}^* constitute a linear base for the forms of degree $\alpha + 1$ which belong to $\mathfrak{A}$. If we multiply the ω_{ij}^* by $\xi_0^*, \cdots, \xi_n^*$, then the products give a base for the forms of degree $\alpha + 2$ in $\mathfrak{A}$. These products are the homogeneous coördinates of the general point of the join V^* of V and $\bar{V}$. Since, as will be pointed out later on, the quadratic transformation has no fundamental points on $\bar{V}$, it follows that $\bar{V}$ and V^* are in regular birational correspondence. We now multiply the homogeneous coördinates of the general point of V^* by $\xi_0^*, \cdots, \xi_n^*$, getting the join of V^* and V. Proceeding in this fashion we construct a model V^*, such that: a) the homogeneous coördinates of the general point of V^* constitute a linear base for the forms of degree $\alpha + a$ in $\mathfrak{A}$; i.e., they are elements of $\mathfrak{A}\mathfrak{p}_0^a$, and consequently also of $\mathfrak{B}$; b) V^* and $\bar{V}$ are in regular birational correspondence. In a similar fashion we construct a model V_1^* in regard to $\mathfrak{B}$, $\bar{V}_1$ and the integer $\beta + b$. Since $\alpha + a = \beta + b$, it follows immediately that the homogeneous coördinates of the general point of V^* are linearly dependent on the homogeneous coördinates of the general point of V_1^*, and *vice versa*. Consequently, V^* and V_1^* are related projectively to each other, q.e.d.

If K is algebraically closed, then it is permissible to assume that the coördinates of P (elements of K) are $1, 0, 0, \cdots, 0$. Then $\mathfrak{p}^* = (\xi_1^*, \cdots, \xi_n^*)$, and our quadratic transformation is given by the equations: $\omega_{ij} = \xi_i \xi_j$, $i = 0, 1, \cdots, n$; $j = 1, 2, \cdots, n$. This is the ordinary quadratic transformation defined by the system of hyperquadrics passing through the point P (see "Reduction," p. 676).

Since P is the only point of V at which all the forms $\xi_i^* \eta_j^*$ vanish simultaneously, P is the only fundamental point of the quadratic transformation. To any other point A of V there corresponds a unique point $\bar{A}$ on $\bar{V}$, and moreover $Q(\bar{A})$ coincides with $Q(A)$. The transformation has no fundamental points on $\bar{V}$. The proofs of all these assertions are straightforward and do not differ essentially from the proofs in the case of an algebraically closed ground field (see "Reduction," p. 676–679).

To see what happens to the point P, we pass to non-homogeneous coördinates: $\xi_i = \xi_i^*/\xi_0^*$ $(i \neq 0)$, and $\omega_{ij} = \omega_{ij}^*/\omega_{00}^*$ $(i, j$ not both zero). Let η_i denote the non-homogeneous polynomial obtained from the form η_i^* by the usual process (replace ξ_0^* by 1 and ξ_i^* by ξ_i, $i \neq 0$). The ring of non-homogeneous coördinates for V, resp. $\bar{V}$, will be: $\mathfrak{o} = \mathsf{K}[\xi_1, \cdots, \xi_n]$ and

$$\bar{\mathfrak{o}} = \mathsf{K}[\xi_1, \cdots, \xi_n, \eta_1/\eta_0, \cdots, \eta_m/\eta_0]$$

respectively. The point P will be given by the prime ideal $\mathfrak{p} = (\eta_0, \cdots, \eta_m)$ in $\mathfrak{o}$. Since the ideal $\bar{\mathfrak{o}}\mathfrak{p}$ is the principal ideal $\bar{\mathfrak{o}} \cdot \eta_0$, we conclude immediately that *the transform $T(P)$ of the point P is a pure $(r - 1)$-dimensional subvariety of $\bar{V}$* (not necessarily irreducible).[10]

[10] If Γ is an irreducible subvariety of $\bar{V}$, at finite distance with respect to the non-homogeneous coördinates ω_{ij}, Γ is given by a prime ideal $\bar{\mathfrak{p}}$ in the ring $\bar{\mathfrak{o}}$. If Γ corresponds to the

4. Quadratic transformations with simple center. The $\mathfrak{p}$-adic divisor

We are now especially interested in the case in which P is a simple point. Let $\mathfrak{J}$ be the quotient ring of P and let $\mathfrak{P}$ denote the ideal of non-units in $\mathfrak{J}$. Let $t_1, \cdots, t_r$, $t_i \in \mathfrak{J}$, be uniformizing parameters of P, i.e. a set of r elements in $\mathfrak{J}$ with the property $\mathfrak{J}\cdot(t_1, \cdots, t_r) = \mathfrak{P}$.[11]

A significant property of the uniformizing parameters is the following: if $\phi(t_1, \cdots, t_r) = 0$, where ϕ is a polynomial in $t_1, \cdots, t_r$ with coefficients in $\mathfrak{J}$, and if $\phi_\rho(t_1, \cdots, t_r)$ is the sum of terms of ϕ of lowest degree ρ (ϕ_ρ = the *leading form* of ϕ), then the coefficients of ϕ_ρ must all be divisible by $\mathfrak{P}$.[12] This property enables us to construct a valuation of Σ in the following fashion. If $\alpha \in \mathfrak{J}$ and if α is exactly divisible by $\mathfrak{P}^\rho$, then $\alpha = \phi_\rho(t_1, \cdots, t_r)$, where ϕ_ρ is a form of degree ρ with coefficients in $\mathfrak{J}$ but not all in $\mathfrak{P}$. Let β be another element of $\mathfrak{J}$, exactly divisible by $\mathfrak{P}^\sigma$, so that $\beta = \psi_\sigma(t_1, \cdots, t_r)$. We have $\alpha\beta \equiv 0(\mathfrak{P}^{\rho+\sigma})$, and since the coefficients of the form $\phi_\rho\psi_\sigma$ are not all in $\mathfrak{P}$, we conclude, by the property of the uniformizing parameters stated above, that $\alpha\beta$ is not divisible by $\mathfrak{P}^{\rho+\sigma+1}$. This shows that if we put $v(\alpha) = \rho$, we get a *discrete valuation* B of Σ, *of rank* 1. We shall call B the $\mathfrak{p}$-*adic divisor of center* P (P — *a* simple point!).

It is not difficult to see that B *is of dimension* $r - 1$, i.e. B is a divisor. For, we have $v(t_i) = 1$, $v(t_i/t_1) = 0$. Were the B-residues of $t_2/t_1, \cdots, t_r/t_1$ algebraically dependent over K, there would exist a form $\phi_\rho(t_1, \cdots, t_m)$, with coefficients in K, not all zero, such that $v(\phi_\rho) > v(t_1^\rho)$. This would imply $\phi_\rho \equiv 0(\mathfrak{P}^{\rho+1})$, which is impossible.

The following theorem puts into evidence the effect of a quadratic transformation of center P in regard to our $\mathfrak{p}$-adic divisor:

THEOREM 1. *If $\bar{V}$ is the transform of V under a quadratic transformation T of simple center P, then the $\mathfrak{p}$-adic divisor of center P is of the first kind with respect to $\bar{V}$, i.e. the center of the divisor on $\bar{V}$ is $(r-1)$-dimensional. This center coincides with the transform $T(P)$ (whence $T(P)$ is irreducible).*

PROOF. We use the notations of the preceding section. We assume that the variety $T(P)$ is not entirely at infinity, i.e. that the principal ideal $\bar{\mathfrak{o}}\cdot\eta_0$ is not unit ideal (see footnote 10).

---

point P, then we must have $\bar{\mathfrak{p}} \bigcap \mathfrak{o} = \mathfrak{p}$ (see footnote 7). Hence $\bar{\mathfrak{o}}\mathfrak{p} \equiv \mathfrak{o}(\mathfrak{p})$, i.e. Γ lies on the subvariety $\overline{W}$ of $\bar{V}$ defined by the principal ideal $\bar{\mathfrak{o}}\cdot\eta_0$. Conversely, if Γ is on $\overline{W}$, then $\bar{\mathfrak{o}}\cdot\mathfrak{p} \equiv 0(\bar{\mathfrak{p}})$, whence $\bar{\mathfrak{p}} \bigcap \mathfrak{o} = 0$ (since $\mathfrak{p}$ is a maximal ideal of $\mathfrak{o}$), and therefore Γ corresponds to P. This shows that $T(P)$ consists of $\overline{W}$ and perhaps of other irreducible components at infinity. Since any irreducible component of $T(P)$ can be assumed to be at finite distance, for a proper choice of the non-homogeneous coördinates (if $\omega^*_{\alpha\beta}$ is different from zero on the given component, we use the quotients $\omega^*_{ij}/\omega^*_{\alpha\beta}$) and since if the principal ideal $\bar{\mathfrak{o}}\cdot\eta_0$ is not the unit ideal, all its isolated prime ideals are $(r-1)$-dimensional, our assertion follows.

[11] See our paper "Algebraic varieties over ground fields of characteristic zero", American Journal of Mathematics, vol. 62 (January, 1940), p. 199.

[12] See loc. cit. in footnote 11, p. 202 and p. 207–208. As a consequence of this property, the quotient ring of a simple point (and also, more generally, the quotient ring of a simple subvariety) is a "p-Reihenring" in the sense of Krull (see Krull, Dimensionstheorie in Stellenringen, Crelle's Journal, vol. 179 (1938)).

To prove the theorem it will be sufficient to prove that the ideal $\overline{\mathfrak{J}}\cdot\eta_0$ is prime, where $\overline{\mathfrak{J}} = \mathfrak{J}\cdot\bar{\mathfrak{o}}$ ($\mathfrak{J}$ = quotient ring of P),[13] and that the irreducible subvariety of $\bar{V}$ defined by this prime ideal is the center of B. Let B' be an arbitrary valuation of center P, whose center on $\bar{V}$ is at finite distance (such valuations exist, since we have assumed that $T(P)$ is not entirely at infinity).

Among the uniformizing parameters $t_1, \cdots, t_r$ of P we take one which has least value in B'. Let it be t_1. Since $\mathfrak{P} = \mathfrak{J}\cdot(\eta_0, \eta_1, \cdots, \eta_m)$, it follows that $t_i \in \mathfrak{J}\cdot(\eta_0, \eta_1, \cdots, \eta_m)$. Consequently $t_i/\eta_0 \in \overline{\mathfrak{J}}$, and, in particular, $t_1/\eta_0 \in \overline{\mathfrak{J}}$. Since the center of B' is at finite distance we conclude that $v_{B'}(t_1) \geqq v_{B'}(\eta_0)$. *This inequality implies that* $\eta_0 \not\equiv 0(\mathfrak{P}^2)$. For, assume that $\eta_0 \equiv 0(\mathfrak{P}^2)$. Then $\eta_0 = \phi_2(t_1, \cdots, t_r)$, where ϕ_2 is a quadratic form in the t's, with coefficients in $\mathfrak{J}$. We can write $\eta_0/t_1 = t_1\phi_2(1, t_2/t_1, \cdots, t_r/t_1)$. By hypothesis, $v_{B'}(t_i/t_1) \geqq 0$. We also have $v_{B'}(t_1) > 0$ and $v_{B'}(\alpha) \geqq 0$, for any element α in $\mathfrak{J}$ (since P is the center of B'). Consequently, $v_{B'}(\eta_0/t_1) > 0$, a contradiction.

Since $\eta_0 \not\equiv 0(\mathfrak{P}^2)$, we have $v_B(\eta_0) = 1$, and since $v_B(\eta_i) \geqq 1$, we conclude that the quotients η_i/η_0 belong to the valuation ring of B. Consequently *the entire ring* $\overline{\mathfrak{J}}$ *is contained in the valuation ring of* B. Therefore the center of B on V is a (irreducible) subvariety $\overline{W}$ of V at finite distance, where $\overline{W}$ will be given by the prime ideal $\overline{\mathfrak{P}}$ in $\overline{\mathfrak{J}}$ consisting of the elements of positive value in B. Now it is clear that $\overline{\mathfrak{J}}\cdot\eta_0 \equiv 0(\overline{\mathfrak{P}})$. On the other hand, let $\bar{\alpha} \in \overline{\mathfrak{J}}$, $v(\bar{\alpha}) > 0$. We can write $\bar{\alpha}$ in the form: $\bar{\alpha} = \phi_\rho(\eta_0, \eta_1, \cdots, \eta_m)/\eta_0^\rho$, where ϕ_ρ is a form of degree ρ, with coefficients in $\mathfrak{J}$. Since $v_B(\bar{\alpha}) > 0$ and $v_B(\eta_0^\rho) = \rho$, we must have $v_B(\phi_\rho) \geqq \rho + 1$. Hence $\phi_\rho = \psi_{\rho+1}(t_1, \cdots, t_r)$, where $\psi_{\rho+1}$ is a form of degree $\rho + 1$ in $t_1, \cdots, t_r$, with coefficients in $\mathfrak{J}$. Since $t_i \equiv 0(\eta_0, \eta_1, \cdots, \eta_m)$, we will also have $\phi_\rho = g_{\rho+1}(\eta_0, \cdots, \eta_m)$, where $g_{\rho+1}$ is again a form of degree $\rho + 1$ in $\eta_0, \cdots, \eta_m$, with coefficients in $\mathfrak{J}$. Consequently, $\bar{\alpha} = g_{\rho+1}/\eta_0^\rho = \eta_0\cdot g_{\rho+1}(1, \eta_1/\eta_0, \cdots, \eta_m/\eta_0)$, whence $\bar{\alpha} \equiv 0(\overline{\mathfrak{J}}\cdot\eta_0)$. We conclude that $\overline{\mathfrak{J}}\cdot\eta_0 = \overline{\mathfrak{P}}$, and this completes the proof.

From the fact that the *irreducible* variety $T(P)$ is defined by a *principal* ideal, namely by the ideal $\bar{\mathfrak{o}}\cdot\eta_0$, it follows $T(P)$ is a *simple* subvariety of $\bar{V}$ (of dimension $r - 1$, and having η_0 as uniformizing parameter). Therefore $T(P)$ contains points which are simple for $\bar{V}$. We shall need, however, the following stronger result:

THEOREM 2. *Every point of* $T(P)$ *is a simple point of* $\bar{V}$.

PROOF. Let $\overline{P}$ be a point of $T(P)$, which we may assume to be a point at finite distance. We consider an arbitrary but fixed valuation B' with centers P and $\overline{P}$ (on V and $\bar{V}$ respectively). Assuming, as we did before, that $v_{B'}(t_1) \leqq v_{B'}(t_i)$, $i = 1, 2, \cdots, r$, we will have $t_i/\eta_0 \in \overline{\mathfrak{J}}$, and since $\mathfrak{J} \subset Q(\overline{P})$, it follows that

$$\frac{t_i}{\eta_0} \in Q(\overline{P}), \qquad i = 1, 2, \cdots, r. \tag{2}$$

[13] Every prime ideal $\bar{\mathfrak{p}}$ of $\bar{\mathfrak{o}}\cdot\eta_0$ contracts to $\mathfrak{p}$ (see footnote 11), and hence $\mathfrak{J} = \mathfrak{o}_\mathfrak{p} \subseteq \bar{\mathfrak{o}}_{\bar{\mathfrak{p}}}$. From this it follows immediately the ideals $\bar{\mathfrak{o}}\cdot\eta_0$ and $\overline{\mathfrak{J}}\cdot\eta_0$ have like decompositions into maximal primary components.

We consider the ring $\mathfrak{o}^* = \mathsf{K}[\xi_1, \cdots, \xi_n, t_2/t_1, \cdots, t_r/t_1]$. Let V^* be the model whose general point (in non-homogeneous coördinates) is $(\xi_1, \cdots, \xi_n, t_2/t_1, \cdots, t_r/t_1)$. Since $v_{B'}(t_i/t_1) \geqq 0$, the center P^* of B' on V^* is a point at finite distance. We will have also the following relations, similar to (2):

$$\frac{\eta_i}{t_1} \epsilon\, Q(P^*), \qquad i = 0, 1, \cdots, m. \tag{3}$$

From (2) and (3) it follows that t_1/η_0 is an element of $Q(\overline{P})$ and that its reciprocal has non-negative value in the valuation B'. Since $\overline{P}$ is the center of B' on $\overline{V}$, we conclude *that* t_1/η_0 *is a unit in* $Q(\overline{P})$. Similarly, t_1/η_0 is a unit in $Q(P^*)$. But then the quotients $t_i/t_1 (= t_i/\eta_0 \cdot \eta_0/t_1)$ are in $Q(\overline{P})$, and the quotients $\eta_i/\eta_0 (= \eta_i/t_1 \cdot t_1/\eta_0)$ are in $Q(P^*)$. Therefore $\bar{\mathfrak{o}} \subset Q(P^*)$ and $\mathfrak{o}^* \subset Q(\overline{P})$, *and this implies*: $Q(P^*) = Q(\overline{P})$. Thus, we have only to show that P^* is a simple point of V^*. For that we have to exhibit uniformizing parameters at P^*.

Let Δ be the residue class field of the point P, i.e. let $\Delta = \mathfrak{o}/\mathfrak{p}$. Here Δ is an algebraic extension of K. Similarly, let Δ^* be the residue class field of P^*. We have $\Delta^* \supseteq \Delta$, since $\mathfrak{o}^* \supseteq \mathfrak{o}$ and since the prime $\mathfrak{o}^*$-ideal $\mathfrak{p}^*$ of the point P^* contracts in $\mathfrak{o}$ to $\mathfrak{p}$. Let $c_1, \cdots, c_n$ be the P-residues of $\xi_1, \cdots, \xi_n$ $(c_i \epsilon\, \Delta)$, and let $d_2, \cdots, d_r$ be the P^*-residues of $t_2/t_1, \cdots, t_r/t_1$ $(d_i \epsilon\, \Delta^*)$. The residue d_i will be the root an irreducible polynomial $f_i(z)$ with coefficients in Δ. Replace each coefficient of $f_i(z)$ by an arbitrary but fixed element of $\mathfrak{o}$ of which it is the residue. We get a certain polynomial $F_i(z)$ with coefficients in $\mathfrak{o}$. We assert that *the r elements*

$$t_1' = t_1, \qquad t_2' = F_2\left(\frac{t_2}{t_1}\right), \cdots, \qquad t_r' = F_r\left(\frac{t_r}{t_1}\right)$$

are uniformizing parameters at P^*, i.e. $t_1', t_2', \cdots, t_r'$ generate in $Q(P^*)$ the ideal of non-units. It is sufficient to show that *the ideal* $\mathfrak{o}^*(t_1', t_2', \cdots, t_r')$ *is the intersection of prime zero-dimensional ideals* (one of these ideals will have to be the ideal defined by the point P^*, since $t_i' = 0$ at P^*). Now this ideal is contained in the prime $(r-1)$-dimensional ideal $\mathfrak{o}^* t_1 (= \mathfrak{o}^* \mathfrak{p}$; the center of the $\mathfrak{p}$-adic divisor on V^*). Modulo this prime ideal the elements $t_2/t_1, \cdots, t_r/t_1$ are algebraically independent, while the residues of $\xi_1, \cdots, \xi_n$ are $c_1, \cdots, c_n$ respectively. Therefore, if we pass to the ring $\mathfrak{o}^*/\mathfrak{o}^* t_1$, we see immediately that the above assertion is equivalent to the following statement: if $z_2, \cdots, z_r$ are indeterminates over Δ, then the polynomials $f_2(z_2), \cdots, f_r(z_r)$ generate in the polynomial ring $\Delta[z_2, \cdots, z_r]$ an ideal which is the intersection of prime zero-dimensional ideals. Since the polynomials $f_i(z_i)$ are irreducible over Δ and *since* Δ *is of characteristic zero*, this statement is true and its proof is straightforward.

II. Resolution of the Singularities of an Algebraic Surface

5. Two lemmas

If W is an arbitrary collection of points on a given irreducible algebraic variety V (for instance, if W is an algebraic subvariety of V), we denote by

$N(W)$ the set of all zero-dimensional valuations B (of the field Σ of rational functions on V) such that the center of B on V is in W.

LEMMA 1. *If W is an algebraic subvariety of V and if for any point P of W there exists a resolving system for $N(P)$, then there also exists a resolving system of $N(W)$.*

PROOF. Let s = dimension W, i.e. s is the highest dimension of the irreducible components of W. For $s = 0$ the lemma is trivial, because W consists then of a finite number of points. We therefore assume that the lemma is true for $s = \rho - 1$ and we prove that it is also true if dim $W = \rho$.

We fix a point P_i on *each* irreducible ρ-dimensional component of W and we consider a resolving system $V_1, \cdots, V_h$ of $N(P_1, P_2, \cdots)$. Let V^* be the join of $V, V_1, \cdots, V_h$. The points of V^* to which there correspond singular points of V_i form an algebraic subvariety W_i^* of V^* $(i = 1, 2, \cdots, h)$. Let W^* be the intersection of $W_1^*, \cdots, W_h^*$. The points of W which correspond to points of W^* form an algebraic subvariety W_1 of W, *of dimension* $< \rho$, since $P_i \in W_1$. It is clear that $(V_1, \cdots, V_h)$ is also a resolving system of $N(W - W_1)$. Hence, if there exists a finite resolving system for $N(W_1)$, this system, together with $V_1, \cdots, V_h$, will give a resolving system for $N(W)$. Since dim $W_1 < \rho$, our induction is complete.

The second lemma deals with algebraic surfaces. Let F be a normal surface, and let P be a point of F. We apply to F a quadratic transformation of center P, getting a surface F_1'. If F_1' is not normal, we pass from F_1' to a corresponding derived normal surface F_1 (see "Results," p. 290). The birational transformation which consists in passing from a given variety to a corresponding derived normal variety shall be referred to in the sequel as a *normal transformation*. We know from section 3 that the transform of P on F_1' is a pure $(r - 1)$-dimensional subvariety of F_1'. We take an arbitrary point P_1' of this subvariety. To P_1' there will correspond on F_1 at most a finite number of points. Let P_1 be one of these points. We now repeat the above procedure, starting with the normal surface F_1 and with the point P_1. We will get first a quadratic transform F_2' of F_1 (with P_1 as center of the quadratic transformation) and then a derived normal variety F_2 of F_2'. On F_2' we select at random a point P_2' which corresponds to P_1, and then we let P_2 be one of the points of F_2 which corresponds to P_2'. In this fashion we proceed indefinitely, getting an infinite sequence of models $F; F_1', F_1; F_2', F_2; \cdots; F_i', F_i; \cdots$, and of points $P; P_1', P_1; P_2', P_2; \cdots; P_i', P_i; \cdots$, where: a) $P_i' \in F_i'$, $P_i \in F_i$; b) the quotient ring $Q(P_i)$ is integrally closed and c) $Q(P) \subset Q(P_1') \subseteq Q(P_1) \subset Q(P_2') \subseteq Q(P_2) \cdots$.

LEMMA 2. *The union of the quotient rings $Q(P_i)$ (or $\lim Q(P_i)$) is the valuation ring of a zero-dimensional valuation of Σ.*

The proof is exactly the same as the one we gave in the case of algebraically closed ground fields ("Reduction," p. 681, Theorem 10).[14]

[14] In that proof we selected an element ζ in $Q(P)$ which is a non-unit of $Q(P)$ but is not in the "tangential" ideal. The consideration of the tangential ideal can be omitted. We know that if $\mathfrak{P}$ is the ideal of non-units in $Q(P)$, $\mathfrak{P} = (\eta_0, \eta_1, \cdots, \eta_m)$, then the ex-

6. The existence of resolving systems

Let F be a normal surface. We shall prove that the hypothesis that the field Σ of rational functions on F does not possess a resolving system leads to a contradiction. Under this hypothesis it follows, in view of Lemma 1, that there must exist a point P on F such that $N(P)$ does not possess a resolving system. By a quadratic transformation of center P, we transform F into a surface F_1' and into the derived normal surface F_1 of F_1'. Let W_1 be the subvariety of F_1 whose points correspond to P. It is clear that $N(P) = N(W_1)$, and hence there must exist a point P_1 on W_1 such that $N(P_1)$ does not possess a resolving system. By repeated application of this argument, we get an infinite sequence of points $P, P_1, \cdots, P_i, \cdots$ of the type considered in the preceding section, such that $N(P_i)$ does not possess a resolving system, $i = 1, 2, \cdots$. Let B be the zero-dimensional valuation whose valuation ring B is the union of the rings $Q(P_i)$. By the local uniformization theorem, let F^* be a model on which the center P^* of B is a simple point. We have $Q(P^*) \subset B$, hence $Q(P^*) \subset Q(P_i)$, for i sufficiently high, say $i \geqq m$. *Every* valuation of center P_i, $i \geqq m$, will have P^* as center on F^*, i.e. $N(P_i) \subseteq N(P^*)$. Therefore F^* *is a resolving surface for* $N(P_i)$, if $i \geqq m$, a contradiction.

7. Proof of the fundamental theorem

Let F, F' be the pair of surfaces which constitute a resolving system for a given set N of zero-dimensional valuations of our field Σ. The birational correspondence between F and F' may have fundamental points on either surface. Our first step consists in the elimination of the fundamental points of one of the two surfaces, say of F, by applying to F a sequence of successive quadratic and normal transformations, as described in the preceding section. As center of the quadratic transformation we take a fundamental point of F, *one at a time*, until we have exhausted all the fundamental points of F. As a result we get some new normal surface, say F_1. If the birational correspondence between F_1 and F' still has fundamental points on F_1, these points must be among the points which correspond to the fundamental points of F. In that case we proceed with F_1 in the same fashion. We assert that *after a finite number of steps we will get a surface* $F_i = \bar{F}$ *such that the birational correspondence between* $\bar{F}$ *and* F' *has no fundamental points on* $\bar{F}$. For, otherwise there would exist an infinite sequence of points $P, P_1, P_2, \cdots, P_i \in F_i$, such that each point P_i is fundamental and such that $Q(P) \subset Q(P_1) \subset \cdots$. Let $B = \lim Q(P_i)$, and let B be the corresponding zero-dimensional valuation (Lemma 2). Let P' be the center of B on F'. Then if i is sufficiently high we will have $Q(P_i) \supseteq Q(P')$, and this is in contradiction with our hypothesis that P_i is a fundamental point of the birational correspondence between F_i and F' (see section 3).

Since under quadratic transformations simple points go into simple points

tended ideal of $\mathfrak{P}$ in $Q(P_1)$ is a principal ideal generated by one of the elements η_i, say by η_0. Take as ζ the element η_0.

(section 4) and since under normal transformations simple points are not affected at all, it follows that the pair of surfaces $(\bar{F}, F')$ is also a resolving system of N.

Our next step is similar to the step just executed, namely we now eliminate by quadratic and normal transformations the fundamental points of F' in the birational correspondence between $\bar{F}$, F'. However, *this time we only eliminate those fundamental points of F' which are simple points of F'.* No singular point of F' will be affected, even if it is a fundamental point. Let $\bar{F}'$ be the transform of F' thus obtained.

We assert that the join F^* of $\bar{F}$ and $\bar{F}'$ is a resolving surface for N.

The proof is straightforward. For, let B be any valuation in the set N. Let $\bar{P}$, P', $\bar{P}'$ and P^* be the centers of B on $\bar{F}$, F', $\bar{F}'$ and F^* respectively. We have to show that P^* is a simple point.

First Case: *P' is a singular point.* Then $\bar{P}$ is simple (since $(\bar{F}, F')$ is a resolving system for N). We have $Q(\bar{P}) \supseteq Q(P')$ (since the birational correspondence between $\bar{F}$ and F' has no fundamental points on $\bar{F}$) and also $Q(P') = Q(\bar{P}')$ (since the singular points of F' have not been affected in the passage from F' to $\bar{F}'$). Therefore $Q(\bar{P}) \supseteq Q(\bar{P}')$ and consequently[15] $Q(\bar{P}) = Q(P^*)$. Since $\bar{P}$ is a simple point, it follows that also P^* is a simple point.

Second Case: *P' is a simple point.* If P' is not a fundamental point of the birational correspondence between $\bar{F}$ and F', then $Q(P') = Q(\bar{P})$ (since $\bar{P}$ is also not a fundamental point), and moreover $Q(P') = Q(\bar{P}')$ (since the points of F' which are not fundamental in the above birational correspondence have not been affected in the passage from F' to $\bar{F}'$). Hence $Q(\bar{P}) = Q(\bar{P}') = Q(P^*)$, and therefore P^* is a simple point. If, on the other hand, P' *is* a fundamental point in the birational correspondence between $\bar{F}$ and F', then $\bar{P}'$ is *not* fundamental for the birational correspondence between $\bar{F}$ and $\bar{F}'$ (because the *simple* fundamental points of F' are eliminated in the course of the passage from F' to $\bar{F}'$). Hence $Q(\bar{P}') \supseteq Q(\bar{P})$, $Q(P^*) = Q(\bar{P}')$. Since P' is simple, also $\bar{P}'$ is simple, and consequently P^* is a simple point, q.e.d.

The Johns Hopkins University

[15] We make use of the following property of the join V^* of two varieties V and V': if P^*, P and P' are corresponding points of V^*, V and V' respectively and if $Q(P) \supseteq Q(P')$, then $Q(P) = Q(P^*)$. The proof is straightforward.

6. The existence of resolving systems

Let F be a normal surface. We shall prove that the hypothesis that the field Σ of rational functions on F does not possess a resolving system leads to a contradiction. Under this hypothesis it follows, in view of Lemma 1, that there must exist a point P on F such that $N(P)$ does not possess a resolving system. By a quadratic transformation of center P, we transform F into a surface F_1' and into the derived normal surface F_1 of F_1'. Let W_1 be the subvariety of F_1 whose points correspond to P. It is clear that $N(P) = N(W_1)$, and hence there must exist a point P_1 on W_1 such that $N(P_1)$ does not possess a resolving system. By repeated application of this argument, we get an infinite sequence of points $P, P_1, \cdots, P_i, \cdots$ of the type considered in the preceding section, such that $N(P_i)$ does not possess a resolving system, $i = 1, 2, \cdots$. Let B be the zero-dimensional valuation whose valuation ring B is the union of the rings $Q(P_i)$. By the local uniformization theorem, let F^* be a model on which the center P^* of B is a simple point. We have $Q(P^*) \subset B$, hence $Q(P^*) \subset Q(P_i)$, for i sufficiently high, say $i \geqq m$. *Every* valuation of center P_i, $i \geqq m$, will have P^* as center on F^*, i.e. $N(P_i) \subseteq N(P^*)$. Therefore *F^* is a resolving surface for* $N(P_i)$, if $i \geqq m$, a contradiction.

7. Proof of the fundamental theorem

Let F, F' be the pair of surfaces which constitute a resolving system for a given set N of zero-dimensional valuations of our field Σ. The birational correspondence between F and F' may have fundamental points on either surface. Our first step consists in the elimination of the fundamental points of one of the two surfaces, say of F, by applying to F a sequence of successive quadratic and normal transformations, as described in the preceding section. As center of the quadratic transformation we take a fundamental point of F, *one at a time*, until we have exhausted all the fundamental points of F. As a result we get some new normal surface, say F_1. If the birational correspondence between F_1 and F' still has fundamental points on F_1, these points must be among the points which correspond to the fundamental points of F. In that case we proceed with F_1 in the same fashion. We assert that *after a finite number of steps we will get a surface $F_i = \bar{F}$ such that the birational correspondence between $\bar{F}$ and F' has no fundamental points on $\bar{F}$.* For, otherwise there would exist an infinite sequence of points $P, P_1, P_2, \cdots, P_i \in F_i$, such that each point P_i is fundamental and such that $Q(P) \subset Q(P_1) \subset \cdots$. Let $B = \lim Q(P_i)$, and let B be the corresponding zero-dimensional valuation (Lemma 2). Let P' be the center of B on F'. Then if i is sufficiently high we will have $Q(P_i) \supseteq Q(P')$, and this is in contradiction with our hypothesis that P_i is a fundamental point of the birational correspondence between F_i and F' (see section 3).

Since under quadratic transformations simple points go into simple points

tended ideal of $\mathfrak{P}$ in $Q(P_1)$ is a principal ideal generated by one of the elements η_i, say by η_0. Take as ζ the element η_0.

(section 4) and since under normal transformations simple points are not affected at all, it follows that the pair of surfaces $(\bar{F}, F')$ is also a resolving system of N.

Our next step is similar to the step just executed, namely we now eliminate by quadratic and normal transformations the fundamental points of F' in the birational correspondence between $\bar{F}$, F'. However, *this time we only eliminate those fundamental points of F' which are simple points of F'.* No singular point of F' will be affected, even if it is a fundamental point. Let $\bar{F}'$ be the transform of F' thus obtained.

We assert that the join F^* of $\bar{F}$ and $\bar{F}'$ is a resolving surface for N.

The proof is straightforward. For, let B be any valuation in the set N. Let $\bar{P}$, P', $\bar{P}'$ and P^* be the centers of B on $\bar{F}$, F', $\bar{F}'$ and F^* respectively. We have to show that P^* is a simple point.

First Case: *P' is a singular point.* Then $\bar{P}$ is simple (since $(\bar{F}, F')$ is a resolving system for N). We have $Q(\bar{P}) \supseteq Q(P')$ (since the birational correspondence between $\bar{F}$ and F' has no fundamental points on $\bar{F}$) and also $Q(P') = Q(\bar{P}')$ (since the singular points of F' have not been affected in the passage from F' to $\bar{F}'$). Therefore $Q(\bar{P}) \supseteq Q(\bar{P}')$ and consequently[15] $Q(\bar{P}) = Q(P^*)$. Since $\bar{P}$ is a simple point, it follows that also P^* is a simple point.

Second Case: *P' is a simple point.* If P' is not a fundamental point of the birational correspondence between $\bar{F}$ and F', then $Q(P') = Q(\bar{P})$ (since $\bar{P}$ is also not a fundamental point), and moreover $Q(P') = Q(\bar{P}')$ (since the points of F' which are not fundamental in the above birational correspondence have not been affected in the passage from F' to $\bar{F}'$). Hence $Q(\bar{P}) = Q(\bar{P}') = Q(P^*)$, and therefore P^* is a simple point. If, on the other hand, P' *is* a fundamental point in the birational correspondence between $\bar{F}$ and F', then $\bar{P}'$ is *not* fundamental for the birational correspondence between $\bar{F}$ and $\bar{F}'$ (because the *simple* fundamental points of F' are eliminated in the course of the passage from F' to $\bar{F}'$). Hence $Q(\bar{P}') \supseteq Q(\bar{P})$, $Q(P^*) = Q(\bar{P}')$. Since P' is simple, also $\bar{P}'$ is simple, and consequently P^* is a simple point, q.e.d.

The Johns Hopkins University

[15] We make use of the following property of the join V^* of two varieties V and V': if P^*, P and P' are corresponding points of V^*, V and V' respectively and if $Q(P) \supseteq Q(P')$, then $Q(P) = Q(P^*)$. The proof is straightforward.

ANNALS OF MATHEMATICS
Vol. 43, No. 3, July, 1942

A NEW HOMOLOGY THEORY. II

BY W. MAYER

(Received March 27, 1942)

1. Euler relations for the new homology groups

We assume familiarity with the definitions and notations of our previous paper, "A new Homology Theory," these Annals, vol. 43, 1942, pp. 370–380.

The homomorphisms

$$C^{i+q} \to C^i, \qquad i = 0, 1, 2, \cdots; 0 < q < p, \tag{1.1}$$

defined by

$$\mathsf{F}^q(\mathsf{K}^{i+q}) = \mathsf{K}^i \tag{1.2}$$

have the group Z_q^{i+q} as nucleus and the group B_{p-q}^i as map.

In this section the underlying simplicial system Σ is supposed to be finite, and if in addition we restrict ourselves to its regular subsystem $\Sigma_{(r)}$, which restriction does not influence the homology groups, the ranks of all groups involved will be finite. Thus (1.1) yields for the ranks

$$r(C^{i+q}) = r(Z_q^{i+q}) + r(B_{p-q}^i), \qquad 0 < q < p, i = 0, 1, 2, \cdots. \tag{1.3}$$

Hence

$$r(C^i) = r(Z_q^i) + r(B_{p-q}^{i-q}) \tag{1.4}$$

is true for $i \geqq q$ and will remain true for $i < q$ *if the rank of a group of negative dimension is put equal to zero.* This follows from

$$C^i = Z_q^i \qquad \text{for } i < q \text{ (and } i = 0, \pm 1, \pm 2, \cdots). \tag{1.5}$$

REMARK: The introduction of chains of negative dimensions here has as its sole purpose to dispense with the handling of sub-cases. (This device was also used in §7 of our previous paper, but differently from the way adopted here.)

Hence we have

$$r(C^i) = r(Z_q^i) + r(B_{p-q}^{i-q}), \qquad 0 < q < p, \quad i = 0, \pm 1, \pm 2, \cdots. \tag{1.6}$$

Furthermore from

$$H_q^i = Z_q^i - B_q^i \tag{1.7}$$

defining the (q, i)-homology groups, we derive

$$r(H_q^i) = r(Z_q^i) - r(B_q^i), \qquad 0 < q < p, \quad i = 0, \pm 1, \pm 2, \cdots. \tag{1.8}$$

Eliminating $r(Z_q^i)$ from (1.6) and (1.8) we arrive at

$$r(H_q^i) = r(C^i) - r(B_q^i) - r(B_{p-q}^{i-q}), \quad 0 < q < p, \qquad i = 0, \pm 1, \pm 2, \cdots. \tag{1.9}$$

Here we replace q and i by $p - q$ and $i - q$ respectively, to get

$$r(H_{p-q}^{i-q}) = r(C^{i-q}) - r(B_{p-q}^{i-q}) - r(B_q^{i-p}), \tag{1.10}$$
$$0 < q < p, \quad i = 0, \pm 1, \pm 2, \cdots.$$

From (1.9) and (1.10) we eliminate $r(B_{p-q}^{i-q})$

$$\begin{cases} r(H_q^i) - r(H_{p-q}^{i-q}) = r(C^i) - r(C^{i-q}) - r(B_q^i) + r(B_q^{i-p}), \\ 0 < q < p, \quad i = 0, \pm 1, \pm 2, \cdots. \end{cases} \tag{1.11}$$

For $i = np + j$, $0 \leqq j < p$ (1.11) becomes

$$r(H_q^{np+j}) - r(H_{p-q}^{np+j-q}) = r(C^{np+j}) - r(C^{np+j-q}) - r(B_q^{np+j}) + r(B_q^{(n-1)p+j}). \tag{1.12}$$

For j and q fixed ($0 \leqq j < p$, $0 < q < p$) we sum the relations (1.12) for $n = 0, 1, 2, \cdots$, to obtain *the Euler relations*

$$\sum_{n=0}^{\infty} [r(H_q^{np+j}) - r(H_{p-q}^{np+j-q})] = \sum_{n=0}^{\infty} [r(C^{np+j}) - r(C^{np+j-q})]. \tag{1.13}$$

The sums appearing in (1.13) are finite since Σ is finite. (For $p = 2, j = q = 1$ the relation (1.13) is the Euler relation for the classical modulo 2 case.)

If we let

$$r_m = r(C^m), \tag{1.14}$$

then r_m is the number of all simplexes of dimension m, simple or not. If on the other hand α_m denotes the number of the simple simplexes of dimension m then r_m is linearly expressible in terms of $\alpha_0, \alpha_1, \cdots, \alpha_m$, with integral coefficients which do not depend on the underlying simplicial system Σ. Consequently the right side of (1.13) is a linear function of $\alpha_0, \alpha_1, \cdots \alpha_N$, where N denotes the dimension of Σ (i.e. N is the largest dimension of a simple simplex of Σ), with coefficients independent of Σ.

For a Euclidean polyhedron $|\Sigma|$, which is a realization of the simplicial system Σ, the left side of (1.13) has been shown to be a topological invariant. Hence this is true for the right side, and being linear in the $\alpha_0, \alpha_1, \cdots \alpha_N$, this side can differ from the Euler number

$$E(\sum) = \alpha_0 - \alpha_1 + \alpha_2 - \cdots + (-1)^N \alpha_N \tag{1.15}$$

only by an integral factor (§4). To determine this factor we may choose the system Σ in the simplest way, and this we do in taking the system Σ_0 composed of one zero-dimensional simplex. Here $\alpha_0 = 1$ and all the other α_i, $i > 0$, are zero. Furthermore for Σ_0

$$r_0 = r_1 = \cdots = r_{p-2} = 1, \qquad r_n = 0, \cdots n > p - 2, \tag{1.16}$$

since only regular chains are admitted. For Σ_0 the right side of (1.13) becomes

$$r_j - r_{j-q} - r_{p+j-q} \tag{1.17}$$

these being the only non-vanishing terms. Two cases arise: a) $j \neq p-1$, and b) $j = p-1$.

Case a, $j = 0, 1, \cdots, p-2$, splits into subcases a_1) $j \geqq q$ and a_2) $j < q$:

a_1) $\quad r_j = 1, \quad r_{j-q} = 1, \quad r_{p+j-q} = 0,$

a_2) $\quad r_j = 1, \quad r_{j-q} = 0, \quad r_{p+j-q} = 1 \cdots j < q-1,$

$0 \cdots j = q-1.$

Hence for $j \neq p-1$ the expression (1.17) *is zero except for $j = q-1$ in which case* (1.17) $= E(\Sigma_0) = 1$.

In Case b, $r_j = 0$, $r_{j-q} = r_{p-q-1} = 1$, $r_{p+j-q} = r_{2p-q-1} = 0$, since $0 < q < p$. *Hence for $j = p-1$ the expression* (1.17) *is* (independent of q) *equal to* $-E(\Sigma_0) = -1$.

Thus the following formula has been proved

$$(1.18)\qquad \begin{cases} \sum_{n=0}^{\infty} [r(H_q^{np+j}) - r(H_{p-q}^{np+j-q})] = f_j^q E(\Sigma), \\ \text{where } f_j^q = \delta_j^{q-1} - \delta_j^{p-1}, \quad \text{with} \quad \delta_i^k = \begin{matrix} 1 \cdots i = k, \\ 0 \cdots i \neq k. \end{matrix} \end{cases}$$

REMARK: $f_j^q = 0$ is possible only for $j \neq q-1$ and $j \neq p-1$ since $q-1 \neq p-1$.

This formula being true for any (finite) Σ suggests first

$$(1.19)\qquad r(H_q^n) = 0, \quad r(H_{p-q}^{n-q}) = 0, \quad \text{for } n \not\equiv -1 \text{ and } n-q \not\equiv -1, \pmod p,$$

and indeed these two relations hold, as will be shown in the next section. It suffices to prove the first of them, since $n \not\equiv -1$ *and* $n - q \not\equiv -1$, (mod p) entails $n - q \not\equiv -1$ *and* $(n-q) - (p-q) \equiv n \not\equiv -1 \pmod p$. In §3 we shall prove

$$(1.20)\qquad r(H_q^{vp+q-1}) \quad \text{and} \quad r(H_q^{vp+p-1})$$

independent of q, likewise suggested by (1.18).

In the following the simplicial system is no longer supposed to be finite.

2. Proof of the rank relations

$$(2.1)\qquad r(H_q^n) = 0 \quad \text{for } n \not\equiv -1 \text{ and } n-q \not\equiv -1, \pmod p.$$

We first prove

(2.2) *For $n \not\equiv -1$,* (mod p) *any (q, n)-cycle is a $(p-q-1)$-boundary.*

(Of course, for $q = p-1$ the above statement is vacuous.)

To prove (2.2) we introduce $p-1$ linear operators D_α, $\alpha = 0, 1, \cdots, p-2$, with

$$(2.3)\qquad D_\alpha(P_1 \cdots P_{n+1}) = (-1)^\alpha \alpha! \sum_{j=1}^{n+1} (P_1 \cdots P_j^{p-\alpha-1} \cdots P_{n+1}),$$

for n-chains, $n \geqq 0$ and with

$$(2.3.1)\qquad D_\alpha \mathsf{K}^n = 0 \quad \text{for } n < 0, \ (\mathsf{K}^n \text{ a chain of negative dimension}).$$

For $n \geqq 0$ and $\alpha = 0, 1, \cdots, p-3$, we derive from (2.3)

$$(2.4)\qquad \begin{cases} \mathsf{F}D_\alpha(P_1 \cdots P_{n+1}) = (-1)^{\alpha+1}(\alpha+1)! \sum_{j=1}^{n+1} (P_1 \cdots P_j^{p-\alpha-2} \cdots P_{n+1}) \\ \qquad\qquad + (-1)^\alpha \alpha! \sum_{j=1}^{n+1} \sum_{k \neq j} (P_1 \cdots P_j^{p-\alpha-1} \cdots P_{n+1})_k . \end{cases}$$

By (2.3), (2.3.1) the last term on the right side of (2.4) equals

$$(-1)^\alpha \alpha! \sum_{k=1}^{n+1} \sum_{j \neq k} (P_1 \cdots P_j^{p-\alpha-1} \cdots P_{n+1})_k = D_\alpha \mathsf{F}(P_1 \cdots P_{n+1}).$$

and therefore we have for $n \geqq 0$ and $\alpha = 0, 1, \cdots, p-3$

$$(2.5)\qquad (\mathsf{F}D_\alpha - D_\alpha \mathsf{F})(P_1 \cdots P_{n+1}) = D_{\alpha+1}(P_1 \cdots P_{n+1}),$$

or, as an operator relation,

$$(2.6)\qquad (\mathsf{F}D_\alpha - D_\alpha \mathsf{F})\mathsf{K}^n = D_{\alpha+1}\mathsf{K}^n$$

which holds for any n (because of (2.3.1.)) and for $\alpha = 0, 1, \cdots, p-3$.

From (2.6) we derive

$$(2.7)\qquad D_\alpha = \sum_{r=0}^{\alpha} (-1)^r \binom{\alpha}{r} \mathsf{F}^{\alpha-r} D_0 \mathsf{F}^r, \quad \alpha = 0, 1, \cdots, p-2.$$

Since

$$(2.8)\qquad D_{p-2}(P_1 \cdots P_{n+1}) = -(p-2)!(n+1)(P_1 \cdots P_{n+1}),$$

we get from (2.7) for $\alpha = p-2$

$$(2.9)\qquad -(p-2)!(n+1)\mathsf{K}^n = \left\{ \sum_{r=0}^{p-2} (-1)^r \binom{p-2}{r} \mathsf{F}^{p-r-2} D_0 \mathsf{F}^r \right\} \mathsf{K}^n.$$

But p being prime,

$$(2.10)\qquad (p-2)! \equiv 1, \qquad \binom{p-2}{r} \equiv (-1)^r (r+1), \pmod p.$$

Hence for any n-dimensional chain K^n, $n = 0, \pm 1, \pm 2, \cdots$ the relation

$$(2.11)\qquad \sum_{r=0}^{p-2} (r+1)\mathsf{F}^{p-r-2} D_0 \mathsf{F}^r \mathsf{K}^n = -(n+1)\mathsf{K}^n$$

holds. For a q-cycle K^n we have $D_0 \mathsf{F}^q \mathsf{K}^n = D_0 \mathsf{F}^{q+1} \mathsf{K}^n = \cdots = 0$, since then either the $\mathsf{F}^q \mathsf{K}^n$, $\mathsf{F}^{q+1}\mathsf{K}^n$, $\cdots$ vanish or they are chains of negative dimensions. The relation (2.11) then becomes

$$-(n+1)\mathsf{K}^n = \mathsf{F}^{p-q-1}\left\{\sum_{r=0}^{q-1}(r+1)\mathsf{F}^{q-r-1}D_0\mathsf{F}^r\mathsf{K}^n\right\} \tag{2.12}$$

thus proving (2.2).

The term corresponding to $r = q-1$ inside the bracket in (2.12) is $qD_0\mathsf{F}^{q-1}\mathsf{K}^n$. We prove (2.1) by showing that for $n - q \not\equiv -1$, (mod p) this term is a boundary:

$$D_0\mathsf{F}^{q-1}\mathsf{K}^n = \mathsf{F}(\ ), \quad \text{for } n - q \not\equiv -1, \text{ (mod } p). \tag{2.13}$$

For $q - 1 > n$ (2.13) is the zero-chain by (2.3.1), and hence a boundary. Here $n \not\equiv q-1$, thus proving (2.13) for $0 > n - q + 1$. The case $n - q + 1 = 0$ can be neglected since by assumption $n - q \not\equiv -1$ (mod p). Thus we may assume $n - q + 1 > 0$ or

$$n \geqq q. \tag{2.14}$$

Let $(P_1 \cdots P_{n+1})$ be a simplex of the q-cycle K^n; then

$$\begin{cases} D_0\mathsf{F}^{q-1}(P_1 \cdots P_{n+1}) = (q-1)!D_0 \sum_{(\alpha_1 \cdots \alpha_{q-1})} (P_1 \cdots P_{n+1})_{\alpha_1 \cdots \alpha_{q-1}} \\ \qquad = (q-1)! \sum_{(\beta_1 \cdots \beta_{n-q+2})} \sum_{j=1}^{n-q+2} (P_{\beta_1} \cdots P_{\beta_j}^{p-1} \cdots P_{\beta_{n-q+2}}), \end{cases} \tag{2.15}$$

the summation running over all the $(n - q + 2)$ faces of $(P_1 \cdots P_{n+1})$. Again we have to introduce linear chain-operators Δ_ρ, $\rho = 1, \cdots, p-2$ which are defined for chains of dimensions $\geqq 1$ by

$$\Delta_\rho(P_1 \cdots P_\nu) = \sum_{[jl]} (P_1 \cdots P_j^{p-\rho} \cdots P_l^{\rho+1} \cdots P_\nu), \qquad \nu \geqq 2, \tag{2.16}$$

and which transform into zero chains of lower dimension. In (2.16) j, l run over $1, 2, \cdots, \nu$, (without repetition, $j \neq l$). Obviously

$$\Delta_\rho = \Delta_\sigma \qquad \text{for } \rho + \sigma = p - 1 \tag{2.17}$$

so that, of the operators Δ_ρ only $\Delta_1, \Delta_2, \cdots \Delta_{(p-1)/2}$, are distinct. Since $n - q + 2 \geqq 2$ we have, similarly to (2.15),

$$\begin{aligned} &\Delta_1\mathsf{F}^{q-1}(P_1 \cdots P_{n+1}) \\ &\qquad = (q-1)! \sum_{(\beta_1 \cdots \beta_{n-q+2})} \sum_{[jl]} (P_{\beta_1} \cdots P_{\beta_j}^{p-1} \cdots P_{\beta_l}^2 \cdots P_{\beta_{n-q+2}}). \end{aligned} \tag{2.18}$$

Here, putting $n - q + 2 = \nu$, we take the boundary

$$\begin{cases} \mathsf{F}\Delta_1\mathsf{F}^{q-1}(P_1 \cdots P_{n+1}) \\ \qquad = -(q-1)! \sum_{(\beta_1 \cdots \beta_\nu)} \sum_{[jl]} (P_{\beta_1} \cdots P_{\beta_j}^{p-2} \cdots P_{\beta_l}^2 \cdots P_{\beta_\nu}) \\ \qquad\quad + 2(\nu-1)(q-1)! \sum_{(\beta_1 \cdots \beta_\nu)} \sum_{j=1}^{\nu} (P_{\beta_1} \cdots P_{\beta_j}^{p-1} \cdots P_{\beta_\nu}) \\ \qquad\quad + (q-1)! \sum_{(\beta_1 \cdots \beta_\nu)} \sum_{[jl]} \sum_{k \neq j,l} (P_{\beta_1} \cdots P_{\beta_j}^{p-1} \cdots P_{\beta_l}^2 \cdots P_{\beta_\nu})_k. \end{cases} \tag{2.19}$$

For $\nu = n - q + 2 \geqq 3$ the last term (2.19) can be written

$$(2.20) \qquad q\cdot(q-1)!\sum_{(\gamma_1\cdots\gamma_{\nu-1})}\sum_{[jl]}(P_{\gamma_1}\cdots P_{\gamma_j}^{p-1}\cdots P_{\gamma_l}^{2}\cdots P_{\gamma_{\nu-1}}),$$

the first sum running over all combinations $(\gamma_1 \cdots \gamma_{\nu-1})$ of $1, 2, \cdots, n+1$. (Such a combination $(\gamma_1 \cdots \gamma_{\nu-1})$ is derived from a combination $(\beta_1 \cdots \beta_\nu)$ in which all of the $\gamma_1, \cdots \gamma_{\nu-1}$ occur. There are $(n+1)-(\nu-1) = (q+\nu-1)-(\nu-1) = q$ such combinations $(\beta_1 \cdots \beta_\nu)$.) But (2.20) $= \Delta_1 \mathsf{F}^q(P_1 \cdots P_{n+1})$. For $\nu = n - q + 2 = 2$ the last term (2.19) vanishes and so does $\Delta_1 \mathsf{F}^q(P_1 \cdots P_{n+1})$, hence (2.19) can be written

$$(2.21) \qquad \begin{cases} (\mathsf{F}\Delta_1 - \Delta_1\mathsf{F})\mathsf{F}^{q-1}(P_1\cdots P_{n+1}) = 2(\nu-1)D_0\mathsf{F}^{q-1}(P_1\cdots P_{n+1}) \\ \qquad\qquad - (q-1)!\sum_{(\beta_1\cdots\beta_\nu)}\sum_{[jl]}(P_{\beta_1}\cdots P_{\beta_j}^{p-2}\cdots P_{\beta_l}^{2}\cdots P_{\beta_\nu}) \end{cases}$$

For $\sigma = 2, 3, \cdots (p+1)/2$ and chains of dimensions $\geqq 1$, we introduce the linear operators

$$(2.22) \qquad \Lambda_\sigma(P_1\cdots P_\nu) = \sum_{[jl]}(P_1\cdots P_j^{p-\sigma}\cdots P_l^{\sigma}\cdots P_\nu), \qquad \nu \geqq 2.$$

(Obviously $\Lambda_{(p-1)/2} = \Lambda_{(q+1)/2}$). Now (2.21) can be written

$$(2.23) \qquad (\mathsf{F}\Delta_1 - \Delta_1\mathsf{F})\mathsf{F}^{q-1}\mathsf{K}^n = 2(\nu-1)D_0\mathsf{F}^{q-1}\mathsf{K}^n - \Lambda_2\mathsf{F}^{q-1}\mathsf{K}^n.$$

Since $\nu - 1 = n - q + 1 \not\equiv 0 \pmod p$ by assumption, (2.13) is true if and only if

$$(2.24) \qquad \Lambda_2\mathsf{F}^{q-1}\mathsf{K}^n = \mathsf{F}(\ ).$$

This follows from (2.23), K^n being a q-cycle.

For $\rho = 2, 3, \cdots (p-1)/2$ we derive equations analogous to (2.19),

$$(2.25) \qquad \begin{cases} \mathsf{F}\Delta_\rho\mathsf{F}^{q-1}(P_1\cdots P_{n+1}) \\ \quad = -\rho(q-1)!\sum_{(\beta_1\cdots\beta_\nu)}\sum_{[jl]}(P_{\beta_1}\cdots P_{\beta_j}^{p-\rho-1}\cdots P_{\beta_l}^{\rho+1}\cdots P_{\beta_\nu}) \\ \qquad + (\rho+1)(q-1)!\sum_{(\beta_1\cdots\beta_\nu)}\sum_{[jl]}(P_{\beta_1}\cdots P_{\beta_j}^{p-\rho}\cdots P_{\beta_l}^{\rho}\cdots P_{\beta_\nu}) \\ \qquad + (q-1)!\sum_{(\beta_1\cdots\beta_\nu)}\sum_{[jl]}\sum_{k\neq j,l}(P_{\beta_1}\cdots P_{\beta_j}^{p-\rho}\cdots P_{\beta_l}^{\rho+1}\cdots P_{\beta_\nu})_k, \end{cases}$$

where the last term on the right is $\Delta_\rho\mathsf{F}^q(P_1 \cdots P_{n+1})$. Hence for the q-cycle K^n we have for $\rho = 2, \cdots, (p-1)/2$.

$$(2.26) \qquad \mathsf{F}\Delta_\rho\mathsf{F}^{q-1}\mathsf{K}^n = -\rho\Delta_{\rho+1}\mathsf{F}^{q-1}\mathsf{K}^n + (\rho+1)\Delta_\rho\mathsf{F}^{q-1}\mathsf{K}^n.$$

But ρ and $\rho + 1$ are $\not\equiv 0$, (mod p), hence from (2.26) it follows that the two chains

$$(2.27) \qquad \Lambda_\rho\mathsf{F}^{q-1}\mathsf{K}^n, \qquad \Lambda_{\rho+1}\mathsf{F}^{q-1}\mathsf{K}^n, \qquad \rho = 2, 3, \cdots, (p-1)/2$$

are simultaneously bounding or not bounding.

For $\rho = (p-1)/2$ the chains (2.27) coincide and (2.26) becomes

$$\mathsf{F}\Delta_{(p-1)/2}\mathsf{F}^{q-1}\mathsf{K}^n = \Lambda_{(p-1)/2}\mathsf{F}^{q-1}\mathsf{K}^n. \tag{2.28}$$

Hence all the $\Lambda_\rho \mathsf{F}^{q-1}\mathsf{K}^n$ bound, $\rho = (p-1)/2, \cdots, 3, 2$, and then (2.13) follows from (2.23), which proves (2.1).

3. Proof of the rank relations

$$r(H_{p-1}^{\nu p+p-2}) = r(H_q^{\nu p+q-1}) \tag{3.1}$$

and

$$r(H_1^{\nu p+p-1}) = r(H_q^{\nu p+p-1}), \quad q = 1, \cdots, p-1, \nu = 0, 1, 2, \cdots. \tag{3.2}$$

We first define a homomorphism

$$H_r^n \to H_s^n, \qquad r < s, \tag{3.3}$$

of the (r, n)-homology group into the (s, n)-homology-group of Σ. Let

$$\{Z^n\}_{B_r^n} \tag{3.4}$$

denote the homology class of the r-cycle Z^n in H_r^n. Then either $r > n$ or $\mathsf{F}^r(Z^n) = 0^{n-r}$, hence either $s > n$ or $\mathsf{F}^s(Z^n) = \mathsf{F}^{s-r}\mathsf{F}^r(Z^n) = 0^{n-s}$ and Z^n is an s-cycle. The mapping of homology classes

$$\{Z^n\}_{B_r^n} \to \{Z^n\}_{B_s^n} \tag{3.5}$$

then defines the homomorphism (3.3).

REMARK: Since a $(p-r)$-boundary of dimension n is a $(p-s)$-boundary of this dimension, $B_r^n \subset B_s^n$, and the mapping (3.5) is independent of the choice of Z^n in the class left in (3.5).

Let us consider the nucleus of the homomorphism (3.3) as defined by (3.5): The class on the left in (3.5) belongs to the nucleus if and only if $Z^n \subset B_s^n$, i.e. if and only if for some chain K^{n+p-s}

$$\mathsf{F}^{p-s}(\mathsf{K}^{n+p-s}) = Z^n. \tag{3.6}$$

For $n \geqq r$ from $\mathsf{F}^r(Z^n) = 0^{n-r}$ then

$$\mathsf{F}^{p+r-s}(\mathsf{K}^{n+p-s}) = 0^{n-r} \tag{3.7}$$

follows. For $n < r$ we have $n + p - s < r + p - s$ and again K^{n+p-s} is an $(r + p - s)$-cycle. *Thus we have proved that* $\{Z^n\}_{B_r^n}$ *belongs to the nucleus of the homomorphism* (3.3) *if and only if a* $(p + r - s)$*-cycle of dimension* $(p + n - s)$ *exists (i.e. an element of* Z_{p+r-s}^{p+n-s}*) such that* (3.6) *holds.*

But (3.6) defines a homomorphism

$$H_{p+r-s}^{p+n-s} \to H_r^n. \tag{3.8}$$

Indeed, Z^n is an r-cycle if K^{n+p-s} is a $(p + r - s)$-cycle. Furthermore, to a $p - (p + r - s) = (s - r)$-boundary K^{n+p-s},

(3.9) $$\mathsf{K}^{n+p-s} = \mathsf{F}^{s-r}(\mathsf{K}^{n+p-r})$$

corresponds in (3.6)

(3.10) $$Z^n = \mathsf{F}^{p-s}\mathsf{F}^{s-r}(\mathsf{K}^{n+p-r}) = \mathsf{F}^{p-r}(\mathsf{K}^{n+p-r}),$$

i.e. a $(p-r)$-boundary. Our above result can be stated as follows:

$\{Z^n\}_{B_r^n}$ belongs to the nucleus of the homomorphism (3.5) if and only if it belongs to the map-group of the homomorphism (3.8). Or: *The nucleus of* (3.5) *is the map of* (3.8).

We now consider the nucleus of the homomorphism (3.8) as defined by (3.6) or by

(3.11) $$\{\mathsf{K}^{p+n-s}\}_{B_{p+r-s}^{p+n-s}} \to \{\mathsf{F}^{p-s}(\mathsf{K}^{p+n-s})\}_{B_r^n}.$$

The class on the left belongs to the nucleus of (3.8) if and only if

(3.12) $$\mathsf{F}^{p-s}(\mathsf{K}^{n+p-s}) = \mathsf{F}^{p-r}(\mathsf{K}_1^{n+p-r})$$

for some chain K_1^{n+p-r}. But then

(3.13) $$\mathsf{F}^{p-s}[\mathsf{K}^{n+p-s} - \mathsf{F}^{s-r}(\mathsf{K}_1^{n+p-r})] = 0^n,$$

and the $(p-s)$-cycle $\mathsf{K}^{n+p-s} - \mathsf{F}^{s-r}(\mathsf{K}_1^{n+p-r})$, since $p-(p+r-s) = s-r$, belongs to the class on the left in (3.11).

Thus the class on the left in (3.11) *belongs to the nucleus of the homomorphism* (3.8) *if and only if this class contains a* $(p-s)$*-cycle.*

If K^{n+p-s} is a $(p-s)$-cycle, then in a way analogous to (3.5) (since $r+p-s > p-s$)

(3.14) $$\{\mathsf{K}^{n+p-s}\}_{B_{p-s}^{n+p-s}} \to \{\mathsf{K}^{n+p-s}\}_{B_{r+p-s}^{n+p-s}}$$

defines a homomorphism

(3.15) $$H_{p-s}^{n+p-s} \to H_{r+p-s}^{n+p-s}.$$

Our last result can now be stated as follows:

The class on the left in (3.11) *belongs to the nucleus of* (3.8) *if and only if it belongs to the map of* (3.15).

Or: *The nucleus of* (3.8) *is the map of* (3.15).

Thus, in the three homomorphisms

(3.16) $$\begin{cases} H_r^n \to H_s^n \\ H_{p+r-s}^{p+n-s} \to H_r^n \\ H_{p-s}^{p+n-s} \to H_{p+r-s}^{p+n-s} \end{cases}$$

as defined by (3.5), (3.11) and (3.14) respectively, the nucleus of the first (second) is the map of the second (third) homomorphism.

Since $p + r - s > p - s$, the relations (3.16) may be continued to give the chain of homomorphisms:

$$(3.17)\qquad \begin{cases} H_r^n \to H_s^n \\ H_{p+r-s}^{n+p-s} \to H_r^n \\ H_{p-s}^{n+p-s} \to H_{p+r-s}^{n+p-s} \\ H_{p-r}^{n+p-r} \to H_{p-s}^{n+p-s} \\ H_{s-r}^{n+p-r} \to H_{p-r}^{n+p-r} \\ H_s^{n+p} \to H_{s-r}^{n+p-r} \\ H_r^{n+p} \to H_s^{n+p} \end{cases}$$

in which the nucleus of any homomorphism coincides with the map of the following one. We write for

$$(3.18)\qquad s = r + 1,\quad n - r = \nu p - 1,\quad \nu = 0, 1, 2, \cdots,\quad r = 1, \cdots, p - 2,$$

the first three homomorphisms (3.17)

$$(3.19)\qquad \begin{aligned} H_r^{\nu p+r-1} &\to H_{r+1}^{\nu p+r-1} \\ H_{p-1}^{\nu p+p-2} &\to H_r^{\nu p+r-1} \\ H_{p-r-1}^{\nu p+p-2} &\to H_{p-1}^{\nu p+p-2} \end{aligned}$$

By our earlier result (2.1) $H_{r+1}^{\nu p+r-1}$ and $H_{p-r-1}^{\nu p+p-2}$ are zero-groups. (Indeed $\nu p + r - 1 \not\equiv -1$ and $\not\equiv r$, (mod p) and $\nu p + p - 2 \not\equiv -1$ and $\not\equiv p - r - 2$, (mod p).) Thus the nucleus of the first homomorphism and hence the map of the second is the group $H_r^{\nu p+r-1}$ itself. Furthermore, the map of the third homomorphism and hence the nucleus of the second one is the zero-subgroup of $H_{p-1}^{\nu p+p-2}$. Therefore the second homomorphism (3.19) is an isomorphism

$$(3.20)\qquad H_{p-1}^{\nu p+p-2} \approx H_r^{\nu p+r-1},$$

thus proving (3.1) (first for $r = 1, 2, \cdots, p - 2$. But for $r = p - 1$ (3.1) becomes an identity).

Now, beginning with the fourth, we write a second triple of homomorphisms (3.17) again using (3.18):

$$(3.21)\qquad \begin{aligned} H_{p-r}^{\nu p+p-1} &\to H_{p-r-1}^{\nu p+p-2} \\ H_1^{\nu p+p-1} &\to H_{p-r}^{\nu p+p-1} \\ H_{r+1}^{(\nu+1)p+r-1} &\to H_1^{\nu p+p-1}. \end{aligned}$$

Here $H_{p-r-1}^{\nu p+p-2}$ and $H_{r+1}^{(\nu+1)p+r-1}$ are zero-groups, and the same reasoning as used before proves the isomorphism

$$(3.22)\qquad H_1^{\nu p+p-1} \approx H_{p-r}^{\nu p+p-1},$$

thus proving (3.2). The ranks (3.1) are in general different from the ranks (3.2), as follows from the Euler relations (1.18) since, in general, $E(\Sigma) \neq 0$.

4. The Euler number

THEOREM (4.1): *If α_i denotes the number of i-dimensional simple simplexes of the finite simplicial system Σ of dimension n then (up to a factor) the Euler number*

$$E(\Sigma) = \sum_{i=0}^{n} (-1)^i \alpha_i$$

is the only invariant of Σ with respect to its subdivisions, which is linear in α_0, $\alpha_1, \cdots, \alpha_n$.

REMARK: In case Σ is a polyhedron, $E(\Sigma)$ (up to a factor) will be the only *topological* invariant linear in $\alpha_1, \cdots, \alpha_n$.

Let $\sum A_k \alpha_k$, $k = 0, 1, \cdots, n$, be an invariant with respect to subdivisions of simplicial systems Σ of dimension n. Then we prove

$$\sum_{k=0}^{n} A_k \alpha_k = fE(\Sigma). \tag{4.2}$$

To prove (4.2) we can choose any system Σ of dimension n since by assumption the A_k do not depend on Σ. We choose for Σ the system composed of the n-simplex $(P_1 \cdots P_{n+1})$ and its faces denoted by Σ_n. Here

$$\alpha_k(\Sigma_n) = \alpha_k^n = \binom{n+1}{k+1}. \tag{4.3}$$

Let Σ_n be subdivided into Σ_n' and let $\alpha_k'^n$ be the number of the k-simplexes of Σ_n'. Then

$$\sum_{k=0}^{n} A_k(\alpha_k'^n - \alpha_k^n) = 0, \tag{4.4}$$

i.e. the "vector" $\xi_k^n = \alpha_k'^n - \alpha_k^n$, $k = 0, 1, \cdots n$, lies in the hyperplane $\sum A_k \xi_k = 0$. If for suitable choices of such subdivisions we arrive at n such linearly-independent vectors $\xi_k^{n,\nu}$, $\nu = 1, 2, \cdots, n$, then the A_k's are determined up to a factor f and (4.2) is proved.

We choose the subdivision Σ_n^ν of Σ_n yielding $\xi_k^{n,\nu}$ $\nu = 1, \cdots n$ (the ν^{th} vector in our construction) by introducing the centroid of a ν-face of Σ_n.

Let $(P_1 \cdots P_{\nu+1})$ be that ν-face, $v = 0$ included, S its centroid and $\alpha_k^{n,\nu}$ the number of k-simplexes of Σ_n^ν. A k-simplex of Σ_n^ν not containing S is of the type $(P_{\alpha_1} \cdots P_{\alpha_{k+1}})$ but not of the type $(P_1 \cdots P_{\nu+1} P_{\beta_1} \cdots P_{\beta_{k-\nu}})$, which (for $v \neq 0$) does not exist in Σ_n^ν.

Hence the number of k-simplexes of Σ_n^ν not containing S is

$$\binom{n+1}{k+1} - \binom{n-\nu}{k-\nu}. \tag{4.5.1}$$

Since $p + r - s > p - s$, the relations (3.16) may be continued to give the chain of homomorphisms:

$$(3.17)\qquad \begin{cases} H_r^n \to H_s^n \\ H_{p+r-s}^{n+p-s} \to H_r^n \\ H_{p-s}^{n+p-s} \to H_{p+r-s}^{n+p-s} \\ H_{p-r}^{n+p-r} \to H_{p-s}^{n+p-s} \\ H_{s-r}^{n+p-r} \to H_{p-r}^{n+p-r} \\ H_s^{n+p} \to H_{s-r}^{n+p-r} \\ H_r^{n+p} \to H_s^{n+p} \end{cases}$$

in which the nucleus of any homomorphism coincides with the map of the following one. We write for

$$(3.18)\qquad s = r + 1, \quad n - r = \nu p - 1, \quad \nu = 0, 1, 2, \cdots, \quad r = 1, \cdots, p - 2,$$

the first three homomorphisms (3.17)

$$(3.19)\qquad \begin{aligned} H_r^{\nu p+r-1} &\to H_{r+1}^{\nu p+r-1} \\ H_{p-1}^{\nu p+p-2} &\to H_r^{\nu p+r-1} \\ H_{p-r-1}^{\nu p+p-2} &\to H_{p-1}^{\nu p+p-2} \end{aligned}$$

By our earlier result (2.1) $H_{r+1}^{\nu p+r-1}$ and $H_{p-r-1}^{\nu p+p-2}$ are zero-groups. (Indeed $\nu p + r - 1 \not\equiv -1$ and $\not\equiv r$, (mod p) and $\nu p + p - 2 \not\equiv -1$ and $\not\equiv p - r - 2$, (mod p).) Thus the nucleus of the first homomorphism and hence the map of the second is the group $H_r^{\nu p+r-1}$ itself. Furthermore, the map of the third homomorphism and hence the nucleus of the second one is the zero-subgroup of $H_{p-1}^{\nu p+p-2}$. Therefore the second homomorphism (3.19) is an isomorphism

$$(3.20)\qquad H_{p-1}^{\nu p+p-2} \approx H_r^{\nu p+r-1},$$

thus proving (3.1) (first for $r = 1, 2, \cdots, p - 2$. But for $r = p - 1$ (3.1) becomes an identity).

Now, beginning with the fourth, we write a second triple of homomorphisms (3.17) again using (3.18):

$$(3.21)\qquad \begin{aligned} H_{p-r}^{\nu p+p-1} &\to H_{p-r-1}^{\nu p+p-2} \\ H_1^{\nu p+p-1} &\to H_{p-r}^{\nu p+p-1} \\ H_{r+1}^{(\nu+1)p+r-1} &\to H_1^{\nu p+p-1}. \end{aligned}$$

Here $H_{p-r-1}^{\nu p+p-2}$ and $H_{r+1}^{(\nu+1)p+r-1}$ are zero-groups, and the same reasoning as used before proves the isomorphism

$$(3.22)\qquad H_1^{\nu p+p-1} \approx H_{p-r}^{\nu p+p-1},$$

thus proving (3.2). The ranks (3.1) are in general different from the ranks (3.2), as follows from the Euler relations (1.18) since, in general, $E(\Sigma) \neq 0$.

4. The Euler number

THEOREM (4.1): *If α_i denotes the number of i-dimensional simple simplexes of the finite simplicial system Σ of dimension n then (up to a factor) the Euler number*

$$E(\Sigma) = \sum_{i=0}^{n} (-1)^i \alpha_i$$

is the only invariant of Σ with respect to its subdivisions, which is linear in $\alpha_0, \alpha_1, \cdots, \alpha_n$.

REMARK: In case Σ is a polyhedron, $E(\Sigma)$ (up to a factor) will be the only *topological* invariant linear in $\alpha_1, \cdots, \alpha_n$.

Let $\sum A_k \alpha_k$, $k = 0, 1, \cdots, n$, be an invariant with respect to subdivisions of simplicial systems Σ of dimension n. Then we prove

$$\sum_{k=0}^{n} A_k \alpha_k = fE(\Sigma). \tag{4.2}$$

To prove (4.2) we can choose any system Σ of dimension n since by assumption the A_k do not depend on Σ. We choose for Σ the system composed of the n-simplex $(P_1 \cdots P_{n+1})$ and its faces denoted by Σ_n. Here

$$\alpha_k(\Sigma_n) = \alpha_k^n = \binom{n+1}{k+1}. \tag{4.3}$$

Let Σ_n be subdivided into Σ_n' and let $\alpha_k'^n$ be the number of the k-simplexes of Σ_n'. Then

$$\sum_{k=0}^{n} A_k(\alpha_k'^n - \alpha_k^n) = 0, \tag{4.4}$$

i.e. the "vector" $\xi_k^n = \alpha_k'^n - \alpha_k^n$, $k = 0, 1, \cdots n$, lies in the hyperplane $\sum A_k \xi_k = 0$. If for suitable choices of such subdivisions we arrive at n such linearly-independent vectors $\xi_k^{n,\nu}$, $\nu = 1, 2, \cdots, n$, then the A_k's are determined up to a factor f and (4.2) is proved.

We choose the subdivision Σ_n^ν of Σ_n yielding $\xi_k^{n,\nu}$ $\nu = 1, \cdots n$ (the ν^{th} vector in our construction) by introducing the centroid of a ν-face of Σ_n.

Let $(P_1 \cdots P_{\nu+1})$ be that ν-face, $v = 0$ included, S its centroid and $\alpha_k^{n,\nu}$ the number of k-simplexes of Σ_n^ν. A k-simplex of Σ_n^ν not containing S is of the type $(P_{\alpha_1} \cdots P_{\alpha_{k+1}})$ but not of the type $(P_1 \cdots P_{\nu+1} P_{\beta_1} \cdots P_{\beta_{k-\nu}})$, which (for $v \neq 0$) does not exist in Σ_n^ν.

Hence the number of k-simplexes of Σ_n^ν not containing S is

$$\binom{n+1}{k+1} - \binom{n-\nu}{k-\nu}. \tag{4.5.1}$$

A k-simplex of Σ_n^ν containing S is of the type $(SP_{\alpha_1} \cdots P\alpha_k)$ but not of the type $(SP_1 \cdots P_{\nu+1}P_{\beta_1} \cdots P_{\beta_{k-\nu-1}})$.
The number of k-simplexes of Σ_n^ν with one vertex S therefore is

$$\binom{n+1}{k} - \binom{n-\nu}{k-\nu-1}. \tag{4.5.2}$$

Thus

$$\alpha_k^{n,\nu} = \binom{n+1}{k+1} + \binom{n+1}{k} - \binom{n-\nu}{k-\nu} - \binom{n-\nu}{k-\nu-1} \tag{4.6}$$

and

$$\xi_k^{n,\nu} = \alpha_k^{n,\nu} - \alpha_k^n = \binom{n+1}{k} - \binom{n-\nu}{k-\nu} - \binom{n-\nu}{k-\nu-1}, \tag{4.7}$$

$$\nu = 0, 1, \cdots, n+1.$$

For $\xi_k^{n,\nu}$ the relations

$$\sum_{k=0}^{n} (-1)^k \xi_k^{n,\nu} = 0 \tag{4.8}$$

hold. This of course is a consequence of the invariance of the Euler number, but is readily verified as follows. The left side of (4.8)

$$= \sum_{k=0}^{n} (-1)^k \binom{n+1}{k} + \sum_{k=1}^{n+1} (-1)^k \binom{n-\nu}{k-\nu-1} - \sum_{k=0}^{n} (-1)^k \binom{n-\nu}{k-\nu-1}$$

$$= (1-1)^{n+1} - (-1)^{n+1} + (-1)^{n+1} = 0.$$

Furthermore,

$$\xi_k^{n,0} = \binom{n+1}{k} - \binom{n}{k} - \binom{n}{k-1} = 0 \tag{4.9}$$

and

$$\xi_0^{n,\nu} = \binom{n+1}{0} - \binom{n-\nu}{-\nu} - \binom{n-\nu}{-\nu-1} = \begin{matrix} 0 \cdots \nu = 0, \\ 1 \cdots \nu \neq 0. \end{matrix} \tag{4.10}$$

From (4.7) for $n, \nu > 0$ we get

$$\begin{aligned} \xi_k^{n-1,\nu-1} &= \binom{n}{k} - \binom{n-\nu}{k-\nu+1} - \binom{n-\nu}{k-\nu} \\ &= \xi_{k+1}^{n,\nu} - \binom{n+1}{k+1} + \binom{n}{k} = \xi_{k+1}^{n,\nu} - \binom{n}{k+1}, \end{aligned} \tag{4.11}$$

or

$$\xi_{k+1}^{n,\nu} = \xi_k^{n-1,\nu-1} + \binom{n}{k+1}. \tag{4.12}$$

We are now able to prove the linear independence of the vectors $\xi_k^{n,\nu}$, $\nu = 1, \cdots, n$, $(k = 0, 1, \cdots, n)$. From

$$\sum_{\nu=1}^{n} a_\nu \xi_k^{n,\nu} = 0 \tag{4.13}$$

by (4.10) for $k = 0$ we deduce

$$\sum_{\nu=1}^{n} a_\nu = 0, \tag{4.14}$$

and then by (4.12) and (4.9) for $k = 1, \cdots, n$

$$\sum_{\nu=1}^{n} a_\nu \xi_{k-1}^{n-1,\nu-1} = \sum_{\nu=0}^{n-1} a_{\nu+1} \xi_{k-1}^{n-1,\nu} = \sum_{\nu=1}^{n-1} a_{\nu+1} \xi_{k-1}^{n-1,\nu} = 0. \tag{4.15}$$

Thus from the linear independence of the $(n - 1)$ vectors $\xi_k^{n-1,\nu}$, $\nu = 1, \cdots, n - 1$ $(k = 0, \cdots, n - 1)$ (by (4.13), (4.14) and (4.15)) the linear independence of the n vectors $\xi_k^{n,\nu}$, $\nu = 1, \cdots, n$, $(k = 0, 1, \cdots, n)$, follows. But $\xi_k^{1,1} = 1$, $k = 0, 1$, hence, of the sets of vectors $\xi_k^{\rho,\nu}$, $\rho = 2, \cdots, n$, $\nu = 1, \cdots, \rho$, $(k = 0, 1, \cdots, \rho)$ each one is a linearly independent set, thus proving (4.1).

INSTITUTE FOR ADVANCED STUDY